ELSEVIER

Bull. Sci. math. 133 (2009) 693–732

BULLETIN DES SCIENCES MATHÉMATIQUES

www.elsevier.com/locate/bulsci

Prime and composite Laurent polynomials ☆

F. Pakovich

Department of Mathematics, Ben Gurion University, P.O.B. 653, Beer Sheva 84105, Israel

Received 18 May 2009

Available online 26 June 2009

Abstract

In the paper [J. Ritt, Prime and composite polynomials, Trans. Amer. Math. Soc. 23 (1922) 51–66] Ritt constructed the theory of functional decompositions of polynomials with complex coefficients. In particular, he described explicitly polynomial solutions of the functional equation $f(p(z)) = g(q(z))$. In this paper we study the equation above in the case where f, g, p, q are holomorphic functions on compact Riemann surfaces. We also construct a self-contained theory of functional decompositions of rational functions with at most two poles generalizing the Ritt theory. In particular, we give new proofs of the theorems of Ritt and of the theorem of Bilu and Tichy.

MSC: primary 30D05; secondary 14H30

Keywords: Ritt's theorems; Decompositions of rational functions; Decompositions of Laurent polynomials

1. Introduction

Let F be a rational function with complex coefficients. The function F is called *indecomposable* if the equality $F = F_2 \circ F_1$, where $F_2 \circ F_1$ denotes a superposition $F_2(F_1(z))$ of rational functions F_1, F_2, implies that at least one of the functions F_1, F_2 is of degree one. Any representation of a rational function F in the form $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$, where $F_1, F_2, \ldots, F_r$ are rational functions, is called *a decomposition* of F. A decomposition is called *maximal* if all $F_1, F_2, \ldots, F_r$ are indecomposable and of degree greater than one.

In general, a rational function may have many maximal decompositions and the ultimate goal of the decomposition theory of rational functions is to describe the general structure of all max-

☆ Research supported by the ISF grant 979/05.
E-mail address: pakovich@math.bgu.ac.il.

0007-4497/$ – see front matter
doi:10.1016/j.bulsci.2009.06.003

imal decompositions up to an equivalence, where by definition two decompositions having an equal number of terms

$$F = F_r \circ F_{r-1} \circ \cdots \circ F_1 \quad \text{and} \quad F = G_r \circ G_{r-1} \circ \cdots \circ G_1$$

are called equivalent if either $r = 1$ and $F_1 = G_1$, or $r \geqslant 2$ and there exist rational functions μ_i, $1 \leqslant i \leqslant r-1$, of degree 1 such that

$$F_r = G_r \circ \mu_{r-1}, \qquad F_i = \mu_i^{-1} \circ G_i \circ \mu_{i-1}, \quad 1 < i < r, \quad \text{and} \quad F_1 = \mu_1^{-1} \circ G_1.$$

Essentially, the unique class of rational functions for which this problem is completely solved is the class of polynomials investigated by Ritt in his classical paper [23].

The results of Ritt can be summarized in the form of two theorems usually called the first and the second Ritt theorems (see [23,26]). The first Ritt theorem states that any two maximal decompositions $\mathcal{D}$, $\mathcal{E}$ of a polynomial P have an equal number of terms and there exists a chain of maximal decompositions $\mathcal{F}_i$, $1 \leqslant i \leqslant s$, of P such that $\mathcal{F}_1 = \mathcal{D}$, $\mathcal{F}_s \sim \mathcal{E}$, and $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$ by replacing two successive functions $A \circ C$ in $\mathcal{F}_i$ by two other functions $B \circ D$ such that

$$A \circ C = B \circ D. \tag{1}$$

The second Ritt theorem states that if A, B, C, D is a polynomial solution of (1) such that

$$\mathrm{GCD}(\deg A, \deg B) = 1, \qquad \mathrm{GCD}(\deg C, \deg D) = 1$$

(this condition is satisfied in particular if A, B, C, D are indecomposable), then there exist polynomials $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$, μ_1, μ_2, where $\deg \mu_1 = 1$, $\deg \mu_2 = 1$, such that

$$A = \mu_1 \circ \tilde{A}, \qquad B = \mu_1 \circ \tilde{B}, \qquad C = \tilde{C} \circ \mu_2, \qquad D = \tilde{D} \circ \mu_2$$

and either

$$\tilde{A} \circ \tilde{C} \sim T_n \circ T_m, \qquad \tilde{B} \circ \tilde{D} \sim T_m \circ T_n,$$

where T_m, T_n are the corresponding Chebyshev polynomials with $n, m \geqslant 1$ and $\mathrm{GCD}(n, m) = 1$, or

$$\tilde{A} \circ \tilde{C} \sim z^n \circ z^r R\big(z^n\big), \qquad \tilde{B} \circ \tilde{D} \sim z^r R^n(z) \circ z^n,$$

where R is a polynomial, $r \geqslant 0$, $n \geqslant 1$, and $\mathrm{GCD}(n, r) = 1$. Actually, the second Ritt theorem essentially remains true for arbitrary polynomial solutions of (1). The only difference in the formulation is that for the degrees of polynomials μ_1, μ_2 in this case the equalities

$$\deg \mu_1 = \mathrm{GCD}(\deg A, \deg B), \qquad \deg \mu_2 = \mathrm{GCD}(\deg C, \deg D)$$

hold (see [6,27]). Notice that an analogue of the second Ritt theorem holds also when the ground field is distinct from $\mathbb{C}$ (see [28]).

For arbitrary rational functions the first Ritt theorem fails to be true. Furthermore, there exist rational functions having maximal decompositions of different length. The simplest examples of such functions can be constructed with the use of rational functions which are Galois coverings. These functions, for the first time calculated by Klein in his famous book [12], are related to the finite subgroups C_n, D_n, A_4, S_4, A_5 of $\mathrm{Aut}\,\mathbb{CP}^1$ and nowadays can be interpreted as Belyi functions of Platonic solids (see [5,14]). Since for such a function f its maximal decompositions correspond to maximal chains of subgroups of its monodromy group G, in order to find maximal decompositions of different length of f it is enough to find the corresponding chains of subgroups of G, and it is not hard to check that for the groups A_4, S_4, and A_5 such chains exist (see e.g. [8]).

The analogues of the second Ritt theorem for arbitrary rational solutions of Eq. (1) are known only in several cases. Let us mention some of them. First, notice that the description of rational solution of (1) under condition that C and D are polynomials turns out to be quite simple and substantially reduces to the description of polynomial solutions of (1) (see [17]). On the other hand, the problem of description of rational solutions of (1) under condition that A and B are polynomials is equivalent to the problem of description of algebraic curves of the form

$$A(x) - B(y) = 0, \tag{2}$$

having a factor of genus zero, together with corresponding parameterizations. A complete list of such curves is known only in the case where the corresponding factor has at most two points at infinity. In this case the problem is closely related to the number theory and was studied first in the paper of Fried [9] and then in the papers of Bilu [2] and Bilu and Tichy [3]. In particular, in [3] an explicit list of such curves, defined over any field of characteristic zero, was obtained. Notice that the results of [9,3] generalize the second Ritt theorem since polynomial solutions of (1) correspond to curves (2) having a factor of genus zero with one point at infinity. Rational solutions of the equation

$$A \circ C = A \circ D, \tag{3}$$

under condition that A is a polynomial were described in [1] (notice also the paper [24] where some partial results about Eq. (3) under condition that A is a rational function were obtained). Finally, a description of permutable rational functions was obtained in [25] (see also [7]). Note that beside of connections with the number theory Eq. (1) has also important connections with different branches of analysis (see e.g. the recent papers [21,17,18,20,22]).

In this paper we study the equation

$$h = f \circ p = g \circ q, \tag{4}$$

where $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ are fixed holomorphic functions on fixed connected compact Riemann surfaces C_1, C_2 and $h : C \to \mathbb{CP}^1$, $p : C \to C_1$, $q : C \to C_2$ are unknown holomorphic functions on unknown connected compact Riemann surface C. We also apply the results obtained to Eq. (1) with rational A, B, C, D and on this base construct a self-contained decomposition theory of rational functions with at most *two* poles generalizing the Ritt theory. In particular, we prove analogues of Ritt theorems for such functions and reprove in a uniform way previous related results of [23,9,2,3].

Let $S \subset \mathbb{CP}^1$ be a finite set and $z_0 \in \mathbb{CP}^1 \setminus S$. Our approach to Eq. (4) is based on the correspondence between pairs consisting of a covering f of $\mathbb{CP}^1$, non-ramified outside of S, together with a point from $f^{-1}\{z_0\}$ and subgroups of finite index in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$. The main advantage of the consideration of such pairs and subgroups, rather than just of functions and their monodromy groups, is due to the fact that for any subgroups of finite index A, B in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ the subgroups $A \cap B$ and $\langle A, B \rangle$ also are subgroups of finite index in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ and we may transfer these operations to the corresponding pairs. The detailed description of the content of the paper is given below.

In Section 2 we describe the general structure of solutions of Eq. (4). We show (Theorem 2.2) that there exists a finite number $o(f, g)$ of solutions h_j, p_j, q_j of (4) such that any other solution may be obtained from them and describe explicitly the monodromy of h_j via the monodromy of f, g. Furthermore, we show (Proposition 2.4) that if f, g are rational functions, then the Riemann surfaces on which the functions h_j, $1 \leqslant j \leqslant o(f,g)$, are defined may be identified with irreducible components of the algebraic curve $f(x) - g(y) = 0$. In particular, being applied to polynomials A, B our construction provides a criterion for the irreducibility of curve (2) via the monodromy groups of A and B useful for applications (see e.g. [20]).

By the analogy with rational functions we will call a pair of holomorphic functions f, g irreducible if $o(f,g) = 1$. In Section 3 we study properties of irreducible and reducible pairs. In particular, we give a criterion (Theorem 3.2) for a pair f, g to be irreducible in terms of the corresponding subgroups of $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ and establish the following result about reducible pairs generalizing the corresponding result of Fried [10] about rational functions (Theorem 3.5): if a pair of holomorphic functions f, g is reducible then there exist holomorphic functions $\tilde{f}$, $\tilde{g}$, p, q such that

$$f = \tilde{f} \circ p, \qquad g = \tilde{g} \circ q, \qquad o(f,g) = o(\tilde{f}, \tilde{g}),$$

and the Galois closures of $\tilde{f}$ and $\tilde{g}$ coincide. We also show (Theorem 3.6) that if in (4) the pair f, g is irreducible then the indecomposability of q implies the indecomposability of f. Notice that the last result turns out to be quite useful for applications related to possible generalizations of the first Ritt theorem (see Section 5).

Further, in Section 4 we study properties of Eq. (4) in the case where f, g are "generalized polynomials" that is holomorphic functions for which the preimage of infinity contains a unique point. In particular, we establish the following, highly useful for the study of Eq. (1), result (Corollary 4.4): if A, B are polynomials of the same degree and C, D are rational functions such that equality (1) holds, then there exist a rational function W, mutually distinct points of the complex sphere γ_i, $1 \leqslant i \leqslant r$, and complex numbers α_i, β_i, $0 \leqslant i \leqslant r$, such that

$$C = \left(\alpha_0 + \frac{\alpha_1}{z - \gamma_1} + \cdots + \frac{\alpha_r}{z - \gamma_r}\right) \circ W, \qquad D = \left(\beta_0 + \frac{\beta_1}{z - \gamma_1} + \cdots + \frac{\beta_r}{z - \gamma_r}\right) \circ W.$$

In Section 5 we propose an approach to possible generalizations of the first Ritt theorem to more wide than polynomials classes of functions. We introduce the conception of a closed class of rational functions as of a subset $\mathcal{R}$ of $\mathbb{C}(z)$ such that the condition $G \circ H \in \mathcal{R}$ implies that $G \in \mathcal{R}$, $H \in \mathcal{R}$. The prototypes for this definition are closed classes $\mathcal{R}_k$, $k \geqslant 1$, consisting of rational functions F for which

$$\min_{z \in \mathbb{CP}^1} \left|F^{-1}\{z\}\right| \leqslant k, \tag{5}$$

where $|F^{-1}\{z\}|$ denotes the cardinality of the set $F^{-1}\{z\}$. Notice that since for any $F \in \mathcal{R}_1$ there exist rational functions μ_1, μ_2 of degree 1 such that $\mu_1 \circ F \circ \mu_2$ is a polynomial, the Ritt theorems can be interpreted as a decomposition theory for the class $\mathcal{R}_1$. The main result of Section 5 (Theorem 5.1) states that in order to check that the first Ritt theorem holds for maximal decompositions of rational functions from a closed class $\mathcal{R}$ it is enough to check that it holds for a certain subset of maximal decompositions which is considerably smaller than the whole set. For example, for the class $\mathcal{R}_1$ this subset turns out to be empty that provides a new proof of the first Ritt for this class (Corollary 5.2). Later, in Section 9, using this method we also show that the first Ritt theorem remains true for the class $\mathcal{R}_2$.

In the rest of the paper, using the results obtained, we construct explicitly the decomposition theory for the class $\mathcal{R}_2$. There are several reasons which make the problem interesting. First, since $\mathcal{R}_1 \subset \mathcal{R}_2$, the decomposition theory for $\mathcal{R}_2$ is a natural generalization of the Ritt theory. Furthermore, the equation

$$L = A \circ C = B \circ D, \tag{6}$$

where $L \in \mathcal{R}_2$ and A, B, C, D are rational functions, is closely related to the equation

$$h = A \circ f = B \circ g, \tag{7}$$

where A, B are rational functions while h, f, g are entire transcendental functions and the description of solutions of (6) yields a description of solutions of (7) (see [22]). Finally, notice that polynomials solutions of (1) naturally appear in the study of the polynomial moment problem which arose recently in connection with the "model" problem for the Poincare center-focus problem (see e.g. [21,4]). The corresponding moment problem for Laurent polynomials, which is related to the Poincare problem even to a greater extent than the polynomial moment problem, is still open and the decomposition theory for $\mathcal{R}_2$ can be considered as a preliminary step in the investigation of this problem.

It was observed by the author several years ago that the description of "double decompositions" (6) of functions from $\mathcal{R}_2$ ("the second Ritt theorem" for $\mathcal{R}_2$) mostly reduces to the classification of curves (2) having a factor of genus 0 with at most two points at infinity. Indeed, without loss of generality we may assume that the minimum in (5) attains at infinity and that $L^{-1}\{\infty\} \subseteq \{0, \infty\}$. In other words, we may assume that L is a Laurent polynomial. Further, it follows easily from the condition $L^{-1}\{\infty\} \subseteq \{0, \infty\}$ that any decomposition $U \circ V$ of L is equivalent either to a decomposition $A \circ L_1$, where A is a polynomial and L_1 is a Laurent polynomial, or to a decomposition $L_2 \circ B$, where L_2 is a Laurent polynomial and $B = cz^d$ for some $c \in \mathbb{C}$ and $d \geqslant 1$. Therefore, the description of double decompositions of functions from $\mathcal{R}_2$ reduces to the solution of the following three equations:

$$A \circ L_1 = B \circ L_2 \tag{8}$$

where A, B are polynomials and L_1, L_2 are Laurent polynomials,

$$A \circ L_1 = L_2 \circ z^d, \tag{9}$$

where A is a polynomial and L_1, L_2 are Laurent polynomials, and

$$L_1 \circ z^{d_1} = L_2 \circ z^{d_2}, \tag{10}$$

where L_1, L_2 are Laurent polynomials. Observe now that if A, B, L_1, L_2 is a solution of Eq. (8), then corresponding curve (2) has a factor of genus 0 with at most two points at infinity and vice versa for any such a curve the corresponding factor may be parametrized by some Laurent polynomials providing a solution of (8). Therefore, the description of solutions of Eq. (8) essentially reduces to the description of curves (2) having a factor of genus 0 with at most two points at infinity. On the other hand, Eqs. (9) and (10) turn out to be much easier for the analysis in view of the presence of symmetries.

Although the result of Bilu and Tichy obtained in the paper [3] (which in its turn uses the results of the papers [2,9,10]) reduces the solution of Eq. (8) to an elementary problem of finding of parameterizations of the corresponding curves, in this paper we give an independent analysis of this equation in view of the following reasons. First, we wanted to provide a self contained exposition of the decomposition theory for the class $\mathcal{R}_2$ since we believe that such an exposition may be interesting for the wide audience. Second, our approach contains some new ideas and by-product results which seem to be interesting by themselves.

Our analysis of Eqs. (8), (9), (10) splits into three parts. In Section 6 using Corollary 4.4 we solve Eqs. (9), (10). In Section 7 using Theorem 3.5 combined with Corollary 4.4 we show (Theorem 7.2) that Eq. (8) in the case where curve (2) is reducible reduces either to the irreducible case or to the case where

$$A \circ L_1 = B \circ L_2 = \frac{1}{2}\left(z^d + \frac{1}{z^d}\right), \quad d > 1.$$

Finally, in Section 8 we solve Eq. (8) in the case where curve (2) is irreducible. Our approach to this case is similar to the one used in the paper [3] and consists of the analysis of the condition that the genus g of (2) is zero. However, we use a different form of the formula for g and replace the conception of "extra" points which goes back to Ritt by a more transparent conception.

Eventually, in Section 9 of the paper, as a corollary of the classification of double decompositions of functions from $\mathcal{R}_2$ and Theorem 5.1, we show (Theorem 9.1) that the first Ritt theorem extends to the class $\mathcal{R}_2$. The results of the paper concerning decompositions of functions from $\mathcal{R}_2$ can be summarized in the form of the following theorem which includes in particular the Ritt theorems and the classifications of curves (2) having a factor of genus 0 with two points at infinity.

Theorem 1.1. *Let*

$$L = A \circ C = B \circ D$$

be two decompositions of a rational function $L \in \mathcal{R}_2$ into compositions of rational functions A, C and B, D. Then there exist rational functions $R, W, \tilde{A}, \tilde{B}, \tilde{C}, \tilde{D} \in \mathcal{R}_2$ such that

$$A = R \circ \tilde{A}, \qquad B = R \circ \tilde{B}, \qquad C = \tilde{C} \circ W, \qquad D = \tilde{D} \circ W, \qquad \tilde{A} \circ \tilde{C} = \tilde{B} \circ \tilde{D}$$

and, up to a possible replacement of A by B and C by D, one of the following conditions holds:

1) $$\tilde{A} \circ \tilde{C} \sim z^n \circ z^r L(z^n), \qquad \tilde{B} \circ \tilde{D} \sim z^r L^n(z) \circ z^n,$$

where L is a Laurent polynomial, $r \geqslant 0$, $n \geqslant 1$, and $\mathrm{GCD}(n, r) = 1$;

2) $$\tilde{A} \circ \tilde{C} \sim z^2 \circ \frac{z^2 - 1}{z^2 + 1} S\left(\frac{2z}{z^2 + 1}\right), \qquad \tilde{B} \circ \tilde{D} \sim \left(1 - z^2\right) S^2(z) \circ \frac{2z}{z^2 + 1},$$

where S is a polynomial;

3) $$\tilde{A} \circ \tilde{C} \sim T_n \circ T_m, \qquad \tilde{B} \circ \tilde{D} \sim T_m \circ T_n,$$

where T_n, T_m are the corresponding Chebyshev polynomials with $m, n \geqslant 1$, and $\mathrm{GCD}(n, m) = 1$;

4) $$\tilde{A} \circ \tilde{C} \sim T_n \circ \frac{1}{2}\left(z^m + \frac{1}{z^m}\right), \qquad \tilde{B} \circ \tilde{D} \sim \frac{1}{2}\left(z^m + \frac{1}{z^m}\right) \circ z^n,$$

where $m, n \geqslant 1$ and $\mathrm{GCD}(n, m) = 1$;

5) $$\tilde{A} \circ \tilde{C} \sim -T_{nl} \circ \frac{1}{2}\left(\varepsilon z^m + \frac{1}{\varepsilon z^m}\right), \qquad \tilde{B} \circ \tilde{D} \sim T_{ml} \circ \frac{1}{2}\left(z^n + \frac{1}{z^n}\right),$$

where T_{nl}, T_{ml} are the corresponding Chebyshev polynomials with $m, n \geqslant 1$, $l > 1$, $\varepsilon^{nl} = -1$, and $\mathrm{GCD}(n, m) = 1$;

6) $$\tilde{A} \circ \tilde{C} \sim \left(z^2 - 1\right)^3 \circ \frac{3(3z^4 + 4z^3 - 6z^2 + 4z - 1)}{(3z^2 - 1)^2},$$
$$\tilde{B} \circ \tilde{D} \sim \left(3z^4 - 4z^3\right) \circ \frac{4(9z^6 - 9z^4 + 18z^3 - 15z^2 + 6z - 1)}{(3z^2 - 1)^3}.$$

Furthermore, if $\mathcal{D}$, $\mathcal{E}$ are two maximal decompositions of L then there exists a chain of maximal decompositions $\mathcal{F}_i$, $1 \leqslant i \leqslant s$, of L such that $\mathcal{F}_1 = \mathcal{D}$, $\mathcal{F}_s \sim \mathcal{E}$, and $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$ by replacing two successive functions in $\mathcal{F}_i$ by two other functions with the same composition.

2. Functional equation $h = f \circ p = g \circ q$

In this section we describe solutions of the functional equation

$$h = f \circ p = g \circ q, \tag{11}$$

where $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ are fixed holomorphic functions on fixed Riemann surfaces C_1, C_2 and $h : C \to \mathbb{CP}^1$, $p : C \to C_1$, $q : C \to C_2$ are unknown holomorphic functions on an unknown Riemann surface C. Notice that substantially we simply describe the components of the fibred product of the covers f and g. Since, however we did not find any exact references to this description in the literature, our exposition is very detailed and essentially self-contained.

We always will assume that the considered Riemann surfaces are connected and compact.

2.1. Preliminaries

Let $S \subset \mathbb{CP}^1$ be a finite set and z_0 be a point from $\mathbb{CP}^1 \setminus S$. Recall that for any collection, consisting of a Riemann surface R, holomorphic function $p : R \to \mathbb{CP}^1$, non-ramified outside of S, and a point $e \in p^{-1}\{z_0\}$, the homomorphism of the fundamental groups

$$p_\star : \pi_1\left(R \setminus p^{-1}\{S\}, e\right) \to \pi_1\left(\mathbb{CP}^1 \setminus S, z_0\right)$$

is a monomorphism such that its image $\Gamma_{p,e}$ is a subgroup of finite index in the group $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$, and vice versa if Γ is a subgroup of finite index in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$, then there exist a Riemann surface R, a function $p : R \to \mathbb{CP}^1$, and a point $e \in p^{-1}\{z_0\}$ such that

$$p_\star\left(\pi_1\left(R \setminus p^{-1}\{S\}, e\right)\right) = \Gamma.$$

Furthermore, this correspondence descends to a one-to-one correspondence between conjugacy classes of subgroups of index d in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ and equivalence classes of holomorphic functions of degree d non-ramified outside of S, where functions $p : R \to \mathbb{CP}^1$ and $\tilde{p} : \tilde{R} \to \mathbb{CP}^1$ are considered as equivalent if there exists an isomorphism $w : R \to \tilde{R}$ such that $p = \tilde{p} \circ w$.

For pairs $p_1 : R_1 \to \mathbb{CP}^1$, $e_1 \in p_1^{-1}\{z_0\}$ and $p_2 : R_2 \to \mathbb{CP}^1$, $e_2 \in p_2^{-1}\{z_0\}$ the groups Γ_{p_1,e_1} and Γ_{p_2,e_2} coincide if and only if there exists an isomorphism $w : R_1 \to R_2$ such that $p_1 = p_2 \circ w$ and $w(e_1) = e_2$. More generally, the inclusion

$$\Gamma_{p_1,e_1} \subseteq \Gamma_{p_2,e_2}$$

holds if and only if there exists a holomorphic function $w : R_1 \to R_2$ such that $p_1 = p_2 \circ w$ and $w(e_1) = e_2$ and in the case if such a function exists it is defined in a unique way. Notice that this implies that if $p: R \to \mathbb{CP}^1$, $e \in p^{-1}\{z_0\}$ is a pair such that

$$\Gamma_{p_1,e_1} \subseteq \Gamma_{p,e} \subseteq \Gamma_{p_2,e_2} \tag{12}$$

and $v: R_1 \to R$, $u: R \to R_2$, are holomorphic function such that $p = p_2 \circ u$, $p_1 = p \circ v$ and $v(e_1) = e$, $u(e) = e_2$ then $w = u \circ v$. In particular, the function w can be decomposed into a composition of holomorphic functions of degree greater than 1 if and only if there exists $\Gamma_{p,e}$ distinct from Γ_{p_1,e_1} and Γ_{p_2,e_2} such (12) holds.

In view of the fact that holomorphic functions can be identified with coverings of Riemann surfaces all the results above follow from the corresponding results about coverings (see e.g. [15]). Notice that the more customary language describing compositions of coverings uses monodromy groups of the functions involved rather than subgroups of $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$. The interaction between these languages is explained below. In the paper we will use both these languages.

Fix a numeration $\{z_1, z_2, \dots, z_r\}$ of points of S and for each i, $1 \leqslant i \leqslant r$, fix a small loop β_i around z_i so that $\beta_1\beta_2\cdots\beta_r = 1$ in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$. If $p: R \to \mathbb{CP}^1$ is a holomorphic function non-ramified outside of S then for each i, $1 \leqslant i \leqslant r$, the loop β_i after the lifting by p induces a permutation $\alpha_i(p)$ of points of $p^{-1}\{z_0\}$. The group G_p generated by $\alpha_i(p)$, $1 \leqslant i \leqslant r$, is called the monodromy group of p. Clearly, the group G_p is transitive and the equality $\alpha_1(p)\alpha_2(p)\cdots\alpha_r(p) = 1$ holds in G_p. The representation of $\alpha_i(p)$, $1 \leqslant i \leqslant r$, by elements of the corresponding symmetric group depends on the numeration of points of $p^{-1}\{z_0\}$ but the conjugacy class of the corresponding collection of permutations is well defined. Moreover, there is a one-to-one correspondence between equivalence classes of holomorphic functions of degree d non-ramified outside of S and conjugacy classes of ordered collections of permutations α_i, $1 \leqslant i \leqslant r$, from the symmetric group S_d acting on the set $\{1, 2, \dots, d\}$ such that $\alpha_1\alpha_2\cdots\alpha_r = 1$ and the permutation group generated by α_i, $1 \leqslant i \leqslant r$, is transitive (see e.g. [16, Corollary 4.10]). We will denote the conjugacy class of permutations which corresponds to a holomorphic function $p: R \to \mathbb{CP}^1$ by $\hat{\alpha}(p)$. If

$$\varphi_p : \pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big) \to G_p \subset S_d$$

is a homomorphism which sends β_i to α_i, $1 \leqslant i \leqslant r$, then the set of preimages of the stabilizers $G_{p,i}$, $1 \leqslant i \leqslant d$, coincides with the set of the groups $\Gamma_{p,e}$, $e \in p^{-1}\{z_0\}$. On the other hand, for any group $\Gamma_{p,e}$, $e \in p^{-1}\{z_0\}$ the collection of permutations α_i, $1 \leqslant i \leqslant r$, induced on the cosets of $\Gamma_{p,e}$ by β_i, $1 \leqslant i \leqslant r$, is a representative of $\hat{\alpha}(p)$.

If a holomorphic function $p: R \to \mathbb{CP}^1$ of degree d can be decomposed into a composition $p = f \circ q$ of holomorphic functions $q: R \to C$ and $f: C \to \mathbb{CP}^1$ then the group G_p has an imprimitivity system Ω_f consisting of $d_1 = \deg f$ blocks such that the collection of permutations of blocks of Ω_f induced by $\alpha_i(p)$, $1 \leqslant i \leqslant r$, is a representative of $\hat{\alpha}(f)$, and vice versa if G_p has an imprimitivity system Ω such that the collection of permutations of blocks of Ω induced by $\alpha_i(p)$, $1 \leqslant i \leqslant r$, is a representative of $\hat{\alpha}(f)$ for some holomorphic function $f: C \to \mathbb{CP}^1$ then there exists a function $q: R \to C$ such that $p = f \circ q$. Notice that if the set $\{1, 2, \dots, d\}$ is identified with the set $p^{-1}\{z_0\}$, then the set of blocks of the imprimitivity system Ω_f corresponding to the decomposition $p = f \circ q$ has the form $\mathcal{B}_i = q^{-1}\{t_i\}$, $1 \leqslant i \leqslant d_1$, where $\{t_1, t_2, \dots, t_{d_1}\} = f^{-1}\{z_0\}$.

If $p = \tilde{f} \circ \tilde{q}$, where $\tilde{f}: \tilde{C} \to \mathbb{CP}^1$, $\tilde{q}: R \to \tilde{C}$, is an other decomposition of p then the imprimitivity systems Ω_f, $\Omega_{\tilde{f}}$ coincide if and only if there exists an automorphism $\mu: \tilde{C} \to C$ such that

$$f = \tilde{f} \circ \mu^{-1}, \qquad q = \mu \circ \tilde{q}.$$

In this case the decompositions $f \circ q$ and $\tilde{f} \circ \tilde{q}$ are called equivalent. Therefore, equivalence classes of decompositions of p are in a one-to-one correspondence with imprimitivity systems of G_p. More generally, if $\mathcal{B}$ is a block of Ω_f and $\mathcal{C}$ is a block of $\Omega_{\tilde{f}}$ such that $\mathcal{B} \cap \mathcal{C}$ is non-empty, then $\mathcal{B}$ and $\mathcal{C}$ have an intersection of cardinality l if and only if there exist holomorphic functions $w : R \to R_1$, $q_1 : R_1 \to C$, $\tilde{q}_1 : R_1 \to \tilde{C}$, where $\deg w = l$, such that

$$q = q_1 \circ w, \qquad \tilde{q} = \tilde{q}_1 \circ w.$$

In particular, if $p = f \circ q = f \circ q_1$ and the imprimitivity systems corresponding to the decompositions $p = f \circ q$ and $p = f \circ q_1$ coincide then $q_1 = \omega \circ q$ where ω is an automorphism of the surface C such that $f \circ \omega = f$. Notice however that in general the equality $f \circ q = f \circ q_1$ does not imply that $q_1 = \omega \circ q$ for some ω as above. On the other hand, since a holomorphic function $q : R \to C$ takes all the values on C the equality $f \circ q = f_1 \circ q$ always implies that $f = f_1$.

By the analogy with rational functions we will call a holomorphic function $p : R \to \mathbb{CP}^1$ of degree greater than 1 indecomposable if the equality $p = f \circ q$ for some holomorphic functions $q : R \to C$ and $f : C \to \mathbb{CP}^1$ implies that at least one of the functions f, q is of degree 1. Clearly, if p is non-ramified outside of S and $z_0 \in \mathbb{CP}^1 \setminus S$, then p is indecomposable if and only if the subgroups $\Gamma_{p,e}$, $e \in p^{-1}\{z_0\}$ are maximal in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$.

2.2. Description of solutions of Eq. (11)

Let $S = \{z_1, z_2, \dots, z_r\}$ be a union of branch points of f, g and z_0 be a fixed point from $\mathbb{CP}^1 \setminus S$.

Proposition 2.1. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be holomorphic functions. Then for any $a \in f^{-1}\{z_0\}$ and $b \in g^{-1}\{z_0\}$ there exist holomorphic functions $u : C \to C_1$, $v : C \to C_2$, $h : C \to \mathbb{CP}^1$, and a point $c \in h^{-1}\{z_0\}$ such that*

$$h = f \circ u = g \circ v, \qquad u(c) = a, \qquad v(c) = b. \tag{13}$$

Furthermore, the function h has the following property: if

$$\tilde{h} = f \circ \tilde{u} = g \circ \tilde{v}, \qquad \tilde{u}(\tilde{c}) = a, \qquad \tilde{v}(\tilde{c}) = b \tag{14}$$

for some holomorphic functions $\tilde{h} : \tilde{C} \to \mathbb{CP}^1$, $\tilde{u} : \tilde{C} \to C_1$, $\tilde{v} : \tilde{C} \to C_2$, and a point $\tilde{c} \in \tilde{h}^{-1}\{z_0\}$, then there exists a holomorphic function $w : \tilde{C} \to C$ such that

$$\tilde{h} = h \circ w, \qquad \tilde{u} = u \circ w, \qquad \tilde{v} = v \circ w, \qquad w(\tilde{c}) = c. \tag{15}$$

Proof. Since the subgroups $\Gamma_{f,a}$ and $\Gamma_{g,b}$ are of finite index in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ their intersection is also of finite index. Therefore, there exists a pair $h : C \to \mathbb{CP}^1$, $c \in h^{-1}\{z_0\}$ such that $\Gamma_{h,c} = \Gamma_{f,a} \cap \Gamma_{g,b}$ and for such a pair equalities (13) hold. Furthermore, equalities (14) imply that $\Gamma_{\tilde{h},\tilde{c}} \subseteq \Gamma_{f,a} \cap \Gamma_{g,b}$ Therefore, $\Gamma_{\tilde{h},\tilde{c}} \subseteq \Gamma_{h,c}$ and hence $\tilde{h} = h \circ w$ for some $w : \tilde{C} \to C$ such that $w(\tilde{c}) = c$. It follows now from

$$f \circ \tilde{u} = f \circ u \circ w, \qquad g \circ \tilde{v} = g \circ v \circ w$$

and

$$(u \circ w)(\tilde{c}) = \tilde{u}(\tilde{c}), \qquad (v \circ w)(\tilde{c}) = \tilde{v}(\tilde{c})$$

that

$$\tilde{u} = u \circ w, \qquad \tilde{v} = v \circ w. \qquad \square$$

For holomorphic functions $f : C_1 \to \mathbb{CP}^1$, $\deg f = n$, and $g : C_2 \to \mathbb{CP}^1$, $\deg g = m$, fix some representatives $\alpha_i(f)$, $\alpha_i(g)$, $1 \leqslant i \leqslant r$, of the classes $\hat{\alpha}(f)$, $\hat{\alpha}(g)$ and define the permutations $\delta_1, \delta_2, \dots, \delta_r \in S_{nm}$ on the set of mn elements c_{j_1,j_2}, $1 \leqslant j_1 \leqslant n$, $1 \leqslant j_2 \leqslant m$, as follows: $c_{j_1,j_2}^{\delta_i} = c_{j_1',j_2'}$, where

$$j_1' = j_1^{\alpha_i(f)}, \qquad j_2' = j_2^{\alpha_i(g)}, \quad 1 \leqslant i \leqslant r.$$

It is convenient to consider c_{j_1,j_2}, $1 \leqslant j_1 \leqslant n$, $1 \leqslant j_2 \leqslant m$, as elements of an $n \times m$ matrix M. Then the action of the permutation δ_i, $1 \leqslant i \leqslant r$, reduces to the permutation of rows of M in accordance with the permutation $\alpha_i(f)$ and the permutation of columns of M in accordance with the permutation $\alpha_i(g)$.

In general the permutation group $\Gamma(f, g)$ generated by δ_i, $1 \leqslant i \leqslant r$, is not transitive on the set c_{j_1,j_2}, $1 \leqslant j_1 \leqslant n$, $1 \leqslant j_2 \leqslant m$. Denote by $o(f, g)$ the number of transitivity sets of $\Gamma(f, g)$ and let $\delta_i(j)$, $1 \leqslant j \leqslant o(f, g)$, $1 \leqslant i \leqslant r$, be the permutation induced by the permutation δ_i, $1 \leqslant i \leqslant r$, on the transitivity set U_j, $1 \leqslant j \leqslant o(f, g)$. By construction, for any j, $1 \leqslant j \leqslant o(f, g)$, the permutation group G_j generated by $\delta_i(j)$, $1 \leqslant i \leqslant r$, is transitive and the equality

$$\delta_1(j)\delta_2(j)\cdots\delta_r(j) = 1$$

holds. Therefore, there exist holomorphic functions $h_j : R_j \to \mathbb{CP}^1$, $1 \leqslant j \leqslant o(f, g)$, such that the collection $\delta_i(j)$, $1 \leqslant i \leqslant r$, is a representative of $\hat{\alpha}(h_j)$. Moreover, it follows from the construction that for each j, $1 \leqslant j \leqslant o(f, g)$, the intersections of the transitivity set U_j with rows of M form an imprimitivity system $\Omega_f(j)$ for G_j such that the permutations of blocks of $\Omega_f(j)$ induced by $\delta_i(j)$, $1 \leqslant i \leqslant r$, coincide with $\alpha_i(f)$. Similarly, the intersections of U_j with columns of M form an imprimitivity system $\Omega_g(j)$ such that the permutations of blocks of $\Omega_g(j)$ induced by $\delta_i(j)$, $1 \leqslant i \leqslant r$, coincide with $\alpha_i(g)$. This implies that there exist holomorphic functions $u_j : R_j \to C_1$ and $v_j : R_j \to C_2$ such that

$$h_j = f \circ u_j = g \circ v_j. \tag{16}$$

Theorem 2.2. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be holomorphic functions. Suppose that $h : R \to \mathbb{CP}^1$, $p : R \to C_1$, $q : R \to C_2$ are holomorphic function such that*

$$h = f \circ p = g \circ q. \tag{17}$$

Then there exist j, $1 \leqslant j \leqslant o(f, g)$, and holomorphic functions $w : R \to R_j$, $\tilde{p} : R_j \to C_1$, $\tilde{q} : R_j \to C_2$ such that

$$h = h_j \circ w, \qquad p = \tilde{p} \circ w, \qquad q = \tilde{q} \circ w \tag{18}$$

and

$$f \circ \tilde{p} \sim f \circ u_j, \qquad g \circ \tilde{q} \sim g \circ v_j.$$

Proof. It follows from Proposition 2.1 that in order to prove the theorem it is enough to show that for any choice of points $a \in f^{-1}\{z_0\}$ and $b \in g^{-1}\{z_0\}$ the class of permutations $\hat{\alpha}(h)$ corresponding to the function h from Proposition 2.1 coincides with $\hat{\alpha}(h_j)$ for some j, $1 \leqslant j \leqslant o(f, g)$. On the other hand, the last statement is equivalent to the statement that for any choice $a \in f^{-1}\{z_0\}$ and $b \in g^{-1}\{z_0\}$ there exist j, $1 \leqslant j \leqslant o(f,g)$, and an element c of the transitivity set U_j such that the group $\Gamma_{f,a} \cap \Gamma_{g,b}$ is the preimage of the stabilizer $G_{j,c}$ of c in the group G_j under the homomorphism

$$\varphi_{h_j} : \pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big) \to G_j$$

(see Section 2.1).

For fixed $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ let l be the index which corresponds to the point a under the identification of the set $f^{-1}\{z_0\}$ with the set $\{1, 2, \dots, n\}$, k be the index which corresponds to the point b under the identification of the set $g^{-1}\{z_0\}$ with the set $\{1, 2, \dots, m\}$, and U_j be the transitivity set of $\Gamma(f,g)$ containing the element $c_{l,k}$. We have

$$\Gamma_{f,a} = \varphi_f^{-1}\{G_{f,l}\}, \qquad \Gamma_{g,b} = \varphi_g^{-1}\{G_{g,k}\}. \tag{19}$$

Furthermore, if $\psi_1 : G_f \to G_j$ (resp. $\psi_2 : G_g \to G_j$) is a homomorphism which sends $\alpha_i(f)$ (resp. $\alpha_i(g)$) to $\alpha_i(h_j)$, $1 \leqslant i \leqslant r$, then

$$G_{f,l} = \psi_1^{-1}\{A_l\}, \qquad G_{g,k} = \psi_2^{-1}\{B_k\}, \tag{20}$$

where A_l (resp. B_k) is the subgroup of G_j which transforms the set of elements $c_{j_1,j_2} \in U_j$ for which $j_1 = a$ (resp. $j_2 = b$) to itself.

Since

$$\psi_1 \circ \varphi_f = \psi_2 \circ \varphi_g = \varphi_{h_j},$$

it follows from (19), (20) that

$$\begin{aligned} \Gamma_{f,a} \cap \Gamma_{g,b} &= (\psi_1 \circ \varphi_f)^{-1}\{A_l\} \cap (\psi_2 \circ \varphi_g)^{-1}\{B_k\} \\ &= \varphi_{h_j}^{-1}\{A_l\} \cap \varphi_{h_j}^{-1}\{B_k\} = \varphi_{h_j}^{-1}\{A_l \cap B_k\} = \varphi_{h_j}^{-1}\{G_{j,c_{k,l}}\}. \qquad \square \end{aligned}$$

For i, $1 \leqslant i \leqslant r$, denote by

$$\lambda_i = (f_{i,1}, f_{i,2}, \dots, f_{i,u_i})$$

the collection of lengths of disjoint cycles in the permutation $\alpha_i(f)$, by

$$\mu_i = (g_{i,1}, g_{i,2}, \dots, g_{i,v_i})$$

the collection of lengths of disjoint cycles in the permutation $\alpha_i(g)$, and by $g(R_j)$, $1 \leqslant j \leqslant o(f,g)$, the genus of the surface R_j. The proposition below generalizes the corresponding result of Fried (see [11, Proposition 2]) concerning the case where f, g are rational functions.

Proposition 2.3. *In the above notation the formula*

$$\sum_{j=1}^{o(f,g)} \big(2 - 2g(R_j)\big) = \sum_{i=1}^{r} \sum_{j_1=1}^{u_i} \sum_{j_2=1}^{v_i} \mathrm{GCD}(f_{i,j_1} g_{i,j_2}) - (r-2)nm \tag{21}$$

holds.

Proof. Denote by $e_i(j)$, $1 \leqslant i \leqslant r$, $1 \leqslant j \leqslant o(f,g)$, the number of disjoint cycles in the permutation $\delta_i(j)$. Since for any j, $1 \leqslant j \leqslant o(f,g)$, the Riemann–Hurwitz formula implies that

$$2 - 2g(R_j) = \sum_{i=1}^{r} e_i(j) - (r-2)|U_j|$$

we have

$$\sum_{j=1}^{o(f,g)} \bigl(2 - 2g(R_j)\bigr) = \sum_{j=1}^{o(f,g)} \sum_{i=1}^{r} e_i(j) - (r-2)mn.$$

On the other hand, it follows from the construction that for given i, $1 \leqslant i \leqslant r$,

$$\sum_{j=1}^{o(f,g)} e_i(j) = \sum_{j_1=1}^{u_i} \sum_{j_2=1}^{v_i} \mathrm{GCD}(f_{i,j_1} g_{i,j_2})$$

and hence

$$\sum_{j=1}^{o(f,g)} \sum_{i=1}^{r} e_i(j) = \sum_{i=1}^{r} \sum_{j_1=1}^{u_i} \sum_{j_2=1}^{v_i} \mathrm{GCD}(f_{i,j_1} g_{i,j_2}). \qquad \square$$

The proposition below shows that if f, g are rational functions then the Riemann surfaces R_j, $1 \leqslant j \leqslant o(f,g)$, may be identified with irreducible components of the affine algebraic curve

$$h_{f,g}(x,y):\ P_1(x)Q_2(y) - P_2(x)Q_1(y) = 0,$$

where P_1, P_2 and Q_1, Q_2 are pairs polynomials without common roots such that

$$f = P_1/P_2, \qquad g = Q_1/Q_2.$$

Proposition 2.4. *For rational functions f, g the corresponding Riemann surfaces R_j, $1 \leqslant j \leqslant o(f,g)$, are in a one-to-one correspondence with irreducible components of the curve $h_{f,g}(x,y)$. Furthermore, each R_j is a desingularization of the corresponding component. In particular, the curve $h_{f,g}(x,y)$ is irreducible if and only if the group $\Gamma(f,g)$ is transitive.*

Proof. For j, $1 \leqslant j \leqslant o(f,g)$, denote by S_j the union of poles of u_j and v_j and define the mapping $t_j : R_j \setminus S_j \to \mathbb{C}^2$ by the formula

$$z \to (u_j, v_j).$$

It follows from formula (16) that for each j, $1 \leqslant j \leqslant o(f,g)$, the mapping t_j maps R_j to an irreducible component of the curve $h_{f,g}(x,y)$. Furthermore, for any point (a,b) on $h_{f,g}(x,y)$, such that $z_0 = f(a) = g(b)$ is not contained in S, there exist uniquely defined j, $1 \leqslant j \leqslant o(f,g)$, and $c \in h_j^{-1}\{z_0\}$ satisfying

$$u_j(c) = a, \qquad v_j(c) = b.$$

This implies that the Riemann surfaces R_j, $1 \leqslant j \leqslant o(f,g)$, are in a one-to-one correspondence with irreducible components of $h_{f,g}(x,y)$ and that each mapping t_j, $1 \leqslant j \leqslant o(f,g)$, is generically injective. Since an injective mapping of Riemann surfaces is an isomorphism onto an open subset we conclude that each R_j is a desingularization of the corresponding component of $h_{f,g}(x,y)$. $\square$

3. Irreducible and reducible pairs

Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be a pair of holomorphic functions non-ramified outside of S and $z_0 \in \mathbb{CP}^1 \setminus S$. By the analogy with the rational case we will call the pair f, g *irreducible* if $o(f, g) = 1$. Otherwise we will call the pair f, g *reducible*. In this section we study properties of irreducible and reducible pairs.

Proposition 3.1. *A pair of holomorphic functions $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ is irreducible whenever their degrees are coprime.*

Proof. Let $n = \deg f$, $m = \deg g$. Since the index of $\Gamma_{f,a} \cap \Gamma_{g,b}$ in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ coincides with the cardinality of the corresponding imprimitivity set U_j, the pair f, g is irreducible if and only if for any $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ the equality

$$\bigl[\pi_1\bigl(\mathbb{CP}^1 \setminus S, z_0\bigr) : \Gamma_{f,a} \cap \Gamma_{g,b}\bigr] = nm \tag{22}$$

holds. Since the index of $\Gamma_{f,a} \cap \Gamma_{g,b}$ in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ is a multiple of the indices of $\Gamma_{f,a}$ and $\Gamma_{g,b}$ in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$, this index is necessary equal to mn whenever n and m are coprime. □

Theorem 3.2. *A pair of holomorphic functions $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ is irreducible if and only if for any $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ the equality*

$$\Gamma_{f,a}\Gamma_{g,b} = \Gamma_{g,b}\Gamma_{f,a} = \pi_1\bigl(\mathbb{CP}^1 \setminus S, z_0\bigr) \tag{23}$$

holds.

Proof. Since

$$\bigl[\pi_1\bigl(\mathbb{CP}^1 \setminus S, z_0\bigr) : \Gamma_{f,a} \cap \Gamma_{g,b}\bigr] = \bigl[\pi_1\bigl(\mathbb{CP}^1 \setminus S, z_0\bigr) : \Gamma_{g,b}\bigr][\Gamma_{g,b} : \Gamma_{f,a} \cap \Gamma_{g,b}],$$

the equality (22) is equivalent to the equality

$$[\Gamma_{g,b} : \Gamma_{f,a} \cap \Gamma_{g,b}] = n. \tag{24}$$

Recall that for any subgroups A, B of finite index in a group G the inequality

$$\bigl[\langle A, B\rangle : A\bigr] \geqslant [B : A \cap B] \tag{25}$$

holds and the equality attains if and only if the groups A and B are permutable (see e.g. [13, p. 79]). Therefore,

$$n = \bigl[\pi_1\bigl(\mathbb{CP}^1 \setminus S, z_0\bigr) : \Gamma_{f,a}\bigr] \geqslant \bigl[\langle \Gamma_{f,a}, \Gamma_{g,b}\rangle : \Gamma_{f,a}\bigr] \geqslant [\Gamma_{g,b} : \Gamma_{f,a} \cap \Gamma_{g,b}]$$

and hence equality (24) holds if and only if $\Gamma_{f,a}$ and $\Gamma_{g,b}$ are permutable and generate $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$. □

Corollary 3.3. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be an irreducible pair of holomorphic functions. Then any pair of holomorphic functions $\tilde{f} : \tilde{C}_1 \to \mathbb{CP}^1$, $\tilde{g} : \tilde{C}_2 \to \mathbb{CP}^1$ such that*

$$f = \tilde{f} \circ p, \qquad g = \tilde{g} \circ q$$

for some holomorphic functions $p : C_1 \to \tilde{C}_1$, $q : C_2 \to \tilde{C}_2$ is also irreducible.

Proof. Since for any $\tilde{a} \in \tilde{f}^{-1}\{z_0\}$, $\tilde{b} \in \tilde{g}^{-1}\{z_0\}$ and $a \in p^{-1}\{\tilde{a}\}$, $b \in q^{-1}\{\tilde{b}\}$ the inclusions

$$\Gamma_{f,a} \subseteq \Gamma_{\tilde{f},\tilde{a}}, \qquad \Gamma_{g,b} \subseteq \Gamma_{\tilde{g},\tilde{b}}$$

hold it follows from (23) that

$$\Gamma_{\tilde{f},\tilde{a}}\Gamma_{\tilde{g},\tilde{b}} = \Gamma_{\tilde{g},\tilde{b}}\Gamma_{\tilde{f},\tilde{a}} = \pi_1(\mathbb{CP}^1 \setminus S, z_0). \qquad \square$$

Set

$$\Gamma_{N_g} = \bigcap_{b \in g^{-1}\{z_0\}} \Gamma_{g,b}$$

and denote by $\hat{N}_g$ the corresponding equivalence class of holomorphic functions. Since the subgroup Γ_{N_g} is normal in $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$, for any $a_1, a_2 \in f^{-1}\{z_0\}$ the subgroups $\Gamma_{f,a_1}\Gamma_{N_g}$ and $\Gamma_{f,a_2}\Gamma_{N_g}$ are conjugated. We will denote the equivalence class of holomorphic functions corresponding to this conjugacy class by $f\hat{N}_g$.

Proposition 3.4. *For any pair of holomorphic functions $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ and a representative $fN_g : C \to \mathbb{CP}^1$ of $f\hat{N}_g$ the equality*

$$o(f, g) = o(fN_g, g)$$

holds.

Proof. For any $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ the action of the permutation group $\Gamma(f, g)$ can be identified with the action of $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ on pairs of cosets $\alpha_{j_1}\Gamma_{f,a}$, $\beta_{j_2}\Gamma_{g,b}$, $1 \leqslant j_1 \leqslant n$, $1 \leqslant j_2 \leqslant m$. Furthermore, two pairs $\alpha_{j_1}\Gamma_{f,a}$, $\beta_{j_2}\Gamma_{g,b}$ and $\alpha_{i_1}\Gamma_{f,a}$, $\beta_{i_2}\Gamma_{g,b}$ are in the same orbit if and only if the set

$$\alpha_{i_1}\Gamma_{f,a}\alpha_{j_1}^{-1} \cap \beta_{i_2}\Gamma_{g,b}\beta_{j_2}^{-1} \tag{26}$$

is non-empty.

Associate now to an orbit $\Gamma(f, g)$ containing the pair $\alpha_{j_1}\Gamma_{f,a}$, $\beta_{j_2}\Gamma_{g,b}$, $1 \leqslant j_1 \leqslant n$, $1 \leqslant j_2 \leqslant m$, an orbit of $\Gamma(fN_g, g)$ containing the pair $\alpha_{j_1}\Gamma_{f,a}\Gamma_{N_g}$, $\beta_{j_2}\Gamma_{g,b}$. If set (26) is non-empty, then the set

$$\alpha_{i_1}\Gamma_{f,a}\Gamma_{N_g}\alpha_{j_1}^{-1} \cap \beta_{i_2}\Gamma_{g,b}\beta_{j_2}^{-1} \tag{27}$$

is also non-empty and therefore we obtain a well-defined map φ from the set of orbits of $\Gamma(f, g)$ to the set of orbits of $\Gamma(fN_g, g)$. Besides, the map φ is clearly surjective.

In order to prove the injectivity of φ we must show that if set (27) is non-empty, then set (26) is also non-empty. So suppose that (27) is non-empty and let x be its element. In view of the normality of Γ_{N_g} the equality

$$\alpha_{i_1}\Gamma_{f,a}\Gamma_{N_g}\alpha_{j_1}^{-1} = \alpha_{i_1}\Gamma_{f,a}\alpha_{j_1}^{-1}\Gamma_{N_g}$$

holds and therefore there exist $\alpha \in \Gamma_{f,a}$, $\beta \in \Gamma_{N_g}$, and $\gamma \in \Gamma_{g,b}$ such that

$$x = \alpha_{i_1}\alpha\alpha_{j_1}^{-1}\beta = \beta_{i_2}\gamma\beta_{j_2}^{-1}.$$

Furthermore, it follows from the definition of Γ_{N_g} that there exists $\gamma_1 \in \Gamma_{g,b}$ such that $\beta_{j_2}^{-1}\beta\beta_{j_2} = \gamma_1$ implying $\beta = \beta_{j_2}\gamma_1\beta_{j_2}^{-1}$. Set $y = x\beta^{-1}$. Then we have

$$y = \alpha_{i_1} \alpha \alpha_{j_1}^{-1} = \beta_{i_2} \gamma \beta_{j_2}^{-1} \beta^{-1} = \beta_{i_2} \gamma \gamma_1^{-1} \beta_{j_2}^{-1}.$$

This implies that y is contained in set (26) and hence (26) is non-empty. □

The following result is a straightforward generalization of the corresponding result of Fried about rational functions (see [10, Proposition 2]).

Theorem 3.5. *For any reducible pair of holomorphic functions* $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ *there exist holomorphic functions* $f_1 : \tilde{C}_1 \to \mathbb{CP}^1$, $g_1 : \tilde{C}_2 \to \mathbb{CP}^1$, *and* $p : C_1 \to \tilde{C}_1$, $q : C_2 \to \tilde{C}_2$ *such that*

$$f = f_1 \circ p, \qquad g = g_1 \circ q, \qquad o(f, g) = o(f_1, g_1), \quad \textit{and} \quad \hat{N}_{f_1} = \hat{N}_{g_1}. \tag{28}$$

Proof. For a holomorphic function $p : R \to \mathbb{CP}^1$ denote by $d(p)$ a maximal number such that there exist holomorphic functions of degree greater than 1

$$p_1 : R \to R_1, \qquad p_i : R_{i-1} \to R_i, \quad 2 \leqslant i \leqslant d(p) - 1, \qquad p_{d(p)} : R_{d(p)-1} \to \mathbb{CP}^1$$

satisfying

$$p = p_{d(p)} \circ p_{d(p)-1} \circ \cdots \circ p_1.$$

We use the induction on the number $d = d(f) + d(g)$.

If $d = 2$ that is if both functions f, g are indecomposable, then the equality $d(f) = 1$, taking into account the normality of N_g, implies that either

$$\Gamma_{f,a} N_g = \Gamma_{f,a} \tag{29}$$

for all $a \in f^{-1}\{z_0\}$ or

$$\Gamma_{f,a} N_g = \pi_1(\mathbb{CP}^1 \setminus S, z_0) \tag{30}$$

for all $a \in f^{-1}\{z_0\}$. The last possibility however would imply that for any $b \in g^{-1}\{z_0\}$

$$\Gamma_{f,a} \Gamma_{g,b} = \Gamma_{g,b} \Gamma_{f,a} = \pi_1(\mathbb{CP}^1 \setminus S, z_0)$$

in contradiction with Theorem 3.2. Therefore, equalities (29) hold and hence

$$N_g \subseteq \bigcap_{a \in f^{-1}\{z_0\}} \Gamma_{f,a} = N_f.$$

The same arguments show that $N_f \subseteq N_g$. Therefore, $N_g = N_f$ and we can set $f_1 = f$, $g_1 = g$.

Suppose now that $d > 2$. If $N_f = N_g$, then as above we can set $f_1 = f$, $g_1 = g$ so assume that $N_f \neq N_g$. Then, again taking into account the normality of N_g, either

$$\Gamma_{f,a} \subsetneq \Gamma_{f,a} N_g \tag{31}$$

for all $a \in f^{-1}\{z_0\}$ or

$$\Gamma_{g,b} \subsetneq \Gamma_{g,b} N_f$$

for all $b \in g^{-1}\{z_0\}$. Suppose say that (31) holds. Since equality (30) is impossible, this implies that for any $a \in f^{-1}\{z_0\}$ there exist $h : C \to \mathbb{CP}^1$ and $c \in h^{-1}\{z_0\}$ such that $\Gamma_{f,a} N_g = \Gamma_{h,c}$.

It follows from (31) that $f = h \circ p$ for some $p : C_1 \to C$ with $1 < \deg h < \deg f$ and hence $d(h) < d(f)$. Since by Proposition 3.4 the equality $o(f, g) = o(h, g)$ holds the theorem follows now from the induction assumption. □

Theorem 3.6. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be an irreducible pair of holomorphic functions and $p : C \to C_1$, $q : C \to C_2$ be holomorphic functions such that $f \circ p = g \circ q$. Suppose that q is indecomposable. Then f is also indecomposable.*

Proof. Set $h = f \circ p = g \circ q$ and fix a point $c \in h^{-1}\{z_0\}$. Since

$$\Gamma_{h,c} \subseteq \Gamma_{f,a}, \qquad \Gamma_{h,c} \subseteq \Gamma_{g,b}, \tag{32}$$

where $a = p(c)$, $b = q(c)$, we have

$$\Gamma_{h,c} \subseteq \Gamma_{f,a} \cap \Gamma_{g,b} \subseteq \Gamma_{g,b}. \tag{33}$$

Furthermore, by Theorem 3.2

$$\Gamma_{f,a}\Gamma_{g,b} = \pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big). \tag{34}$$

Since (34) implies that $\Gamma_{f,a} \cap \Gamma_{g,b} \neq \Gamma_{g,b}$ it follows from (33) taking into account the indecomposability of q that

$$\Gamma_{h,c} = \Gamma_{f,a} \cap \Gamma_{g,b}. \tag{35}$$

In order to prove the theorem we must show that if $\Gamma \subseteq \pi_1(\mathbb{CP}^1 \setminus S, z_0)$ is a subgroup such that

$$\Gamma_{f,a} \subsetneq \Gamma \tag{36}$$

then $\Gamma = \pi_1(\mathbb{CP}^1 \setminus S, z_0)$. Clearly, (34) implies that

$$\Gamma\Gamma_{g,b} = \pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big). \tag{37}$$

Consider the intersection

$$\Gamma_1 = \Gamma \cap \Gamma_{g,b}.$$

It follows from (25) and (34), (37) that

$$\big[\pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big) : \Gamma_{f,a}\big] = [\Gamma_{g,b} : \Gamma_{h,c}], \qquad \big[\pi_1\big(\mathbb{CP}^1 \setminus S, z_0\big) : \Gamma\big] = [\Gamma_{g,b} : \Gamma_1].$$

Therefore, (36) implies that

$$[\Gamma_{g,b} : \Gamma_1] < [\Gamma_{g,b} : \Gamma_{h,c}]$$

and hence $\Gamma_{h,c} \subsetneq \Gamma_1$. Since $\Gamma_1 \subseteq \Gamma_{g,b}$ it follows now from the indecomposability of q that $\Gamma_1 = \Gamma_{g,b}$. Therefore, $\Gamma_{g,b} \subseteq \Gamma$. Since also $\Gamma_{f,a} \subseteq \Gamma$ it follows now from (34) that $\Gamma = \pi_1(\mathbb{CP}^1 \setminus S, z_0)$. $\square$

4. Double decompositions involving generalized polynomials

Say that a holomorphic function $h : C \to \mathbb{CP}^1$ is *a generalized polynomial* if $h^{-1}\{\infty\}$ consists of a unique point. In this section we mention some specific properties of double decompositions $f \circ p = g \circ q$ in the case when f, g are generalized polynomials.

We start from mentioning two corollaries of Theorem 3.5 for such double decompositions.

Corollary 4.1. *If in Theorem* 3.5 *the functions f, g are generalized polynomials then* $\deg f_1 = \deg g_1$.

Proof. The equality $f = f_1 \circ p$ for a generalized polynomial f implies that f_1 is also a generalized polynomial. Furthermore, since $\Gamma_{N_{f_1}} = \bigcap_{a \in f_1^{-1}\{z_0\}} \Gamma_{f_1,a}$ the monodromy group of $\Gamma_{N_{f_1}}$ may be obtained by the repeated use of the construction given in Section 2.2. On the other hand, it is easy to see that if f_1 is a generalized polynomial then on each stage of this process the permutation corresponding to the loop around infinity consists of cycles of length equal to the degree of f_1 only. Therefore, the same is true for $\Gamma_{N_{f_1}}$ and hence the equality $\hat{N}_{f_1} = \hat{N}_{g_1}$ implies that $\deg f_1 = \deg g_1$. □

The following important specification of Theorem 3.5 goes back to Fried (see [10, Proposition 2]).

Corollary 4.2. *Let A, B be polynomials such that curve* (2) *is reducible. Then there exist polynomials A_1, B_1, C, D such that*

$$A = A_1 \circ C, \qquad B = B_1 \circ D, \qquad \hat{N}_{A_1} = \hat{N}_{B_1}, \tag{38}$$

and each irreducible component $F(x, y)$ of curve (2) *has the form $F_1(C(x), D(y))$, where $F_1(x, y)$ is an irreducible component of the curve*

$$A_1(x) - B_1(y) = 0. \tag{39}$$

Proof. Indeed, it follows from Theorem 3.5 and Proposition 2.4 that there exist polynomials A_1, B_1, C, D such that equalities (38) hold and curves (2) and (39) have the same number of irreducible components. Since for each irreducible component $F_1(x, y)$ of curve (39) the polynomial $F_1(C(x), D(y))$ is a component of curve (2), this implies that any irreducible component $F(x, y)$ of curve (2) has the form $F_1(C(x), D(y))$ for some irreducible component $F_1(x, y)$ of curve (39). □

For a holomorphic function $h : C \to \mathbb{CP}^1$ and $z \in C$ denote by $\mathrm{mult}_z\, h$ the multiplicity of h at z.

Theorem 4.3. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be generalized polynomials,* $\deg f = n$, $\deg g = m$, $l = \mathrm{LCM}(n, m)$, *and $h : R \to \mathbb{CP}^1$, $p : R \to C_1$, $q : R \to C_2$ be holomorphic functions such that*

$$h = f \circ p = g \circ q. \tag{40}$$

Then there exist holomorphic functions $w : R \to C$, $\tilde{p} : C \to C_1$, $\tilde{q} : C \to C_2$ such that

$$p = \tilde{p} \circ w, \qquad q = \tilde{q} \circ w, \tag{41}$$

and for any $z \in h^{-1}\{\infty\}$

$$\mathrm{mult}_z\, \tilde{p} = l/n, \qquad \mathrm{mult}_z\, \tilde{q} = l/m.$$

Proof. In view of Theorem 2.2 it is enough to prove that if u_j, v_j, $1 \leqslant j \leqslant o(f, g)$, are functions defined in Section 2.2 then for any $z \in h^{-1}\{\infty\}$ and j, $1 \leqslant j \leqslant o(f, g)$, the equalities

$$\mathrm{mult}_z\, u_j = l/n, \qquad \mathrm{mult}_z\, v_j = l/m \tag{42}$$

hold.

Since f, g are generalized polynomials it follows from the construction given in Section 2.2 that for any function $h_j = f \circ u_j = g \circ v_j$, $1 \leqslant j \leqslant o(f,g)$, the permutation of its monodromy group corresponding to the loop around infinity consists of cycles of length equal to l only. On the other hand, the length of such a cycle coincides with the multiplicity of the corresponding point from $h_j^{-1}\{\infty\}$. Now equalities (42) follow from the fact that for any $z \in R_j$, $1 \leqslant j \leqslant o(f,g)$,

$$\mathrm{mult}_z h_j = \mathrm{mult}_{u_j(z)} f \,\mathrm{mult}_z u_j = \mathrm{mult}_{v_j(z)} g \,\mathrm{mult}_z v_j. \qquad \square$$

Corollary 4.4. *Let A, B be polynomials of the same degree n and C, D be rational functions such that*

$$A \circ C = B \circ D.$$

Then there exist a rational function W, mutually distinct points of the complex sphere γ_i, $1 \leqslant i \leqslant r$, and complex numbers α_i, β_i, $0 \leqslant i \leqslant r$, such that

$$C = \left(\alpha_0 + \frac{\alpha_1}{z-\gamma_1} + \cdots + \frac{\alpha_r}{z-\gamma_r}\right) \circ W, \qquad D = \left(\beta_0 + \frac{\beta_1}{z-\gamma_1} + \cdots + \frac{\beta_r}{z-\gamma_r}\right) \circ W.$$

Furthermore, if α is the leading coefficient of A and β is the leading coefficient of B then $\alpha\alpha_i^n = \beta\beta_i^n$, $1 \leqslant i \leqslant r$.

Proof. Since $\deg A = \deg B$ it follows from Theorem 4.3 that there exist rational functions A, B, W such that $C = \tilde{C} \circ W$, $D = \tilde{D} \circ W$, and all the poles of $\tilde{C}$ and $\tilde{D}$ are simple (the functions $\tilde{C}$ and $\tilde{D}$ obviously have the same set of poles coinciding with the set of poles of the function $A \circ \tilde{C} = B \circ \tilde{D}$). Denoting these poles by γ_i, $1 \leqslant i \leqslant r$, we conclude that

$$\tilde{C} = \alpha_0 + \frac{\alpha_1}{z-\gamma_1} + \cdots + \frac{\alpha_r}{z-\gamma_r}, \qquad \tilde{D} = \beta_0 + \frac{\beta_1}{z-\gamma_1} + \cdots + \frac{\beta_r}{z-\gamma_r}$$

for some $\alpha_i, \beta_i \in \mathbb{C}$, $0 \leqslant i \leqslant r$ (in case if $\gamma_i = \infty$ for some i, $1 \leqslant i \leqslant r$, the corresponding terms should be changed to $\alpha_i z$, $\beta_i z$).

Furthermore, if α (resp. β) is the leading coefficient of A (resp. B) then the leading coefficient of the Laurent expansion of the function $A \circ \tilde{C}$ (resp. $B \circ \tilde{D}$) near γ_i, $1 \leqslant i \leqslant r$, equals $\alpha\alpha_i^n$ (resp. $\beta\beta_i^n$). Since $A \circ \tilde{C} = B \circ \tilde{D}$ this implies that for any i, $1 \leqslant i \leqslant r$, the equality $\alpha\alpha_i^n = \beta\beta_i^n$ holds. $\square$

Notice that replacing the rational function W in Corollary 4.4 by the function $\mu \circ W$, where μ is an appropriate automorphism of the sphere, we may assume that $\gamma_1, \gamma_2, \gamma_3$ are any desired points of the sphere.

Finally, let us mention the following corollary of Theorem 4.3 which generalizes the corresponding property of polynomial decompositions.

Corollary 4.5. *Suppose that under assumptions of Theorem* 4.3 *the function h is a generalized polynomial and* $\deg f = \deg g$. *Then $f \circ p \sim g \circ q$.*

Proof. Set $x = f^{-1}\{\infty\}$. The conditions of the corollary and Theorem 4.3 imply that $\tilde{p}^{-1}\{x\}$ contains a unique point and the multiplicity of this point with respect to $\tilde{p}$ is one. Therefore $\tilde{p}$ is an automorphism. The same is true for $\tilde{q}$. $\square$

5. Ritt classes of rational functions

As it was mentioned above the first Ritt theorem fails to be true for arbitrary rational functions and it is quite interesting to describe the classes of rational functions for which this theorem remains true. In this section we propose an approach to this problem. This approach is especially useful when a sufficiently complete information about double decompositions of the functions from the corresponding class is available. In particular, our method permits to generalize the first Ritt theorem to Laurent polynomials using the classification of their double decompositions.

It is natural to assume that considered classes of rational functions possess some property of closeness which is formalized in the following definition. Say that a set of rational functions $\mathcal{R}$ is *a closed class* if for any $F \in \mathcal{R}$ the equality $F = G \circ H$ implies that $G \in \mathcal{R}$, $H \in \mathcal{R}$. For example, rational functions for which

$$\min_{z \in \mathbb{CP}^1} \left| F^{-1}\{z\} \right| \leqslant k,$$

where $k \geqslant 1$ is a fixed number and $|F^{-1}\{z\}|$ denotes the cardinality of the set $F^{-1}\{z\}$, form a closed class. We will denote this class by $\mathcal{R}_k$.

Say that two maximal decompositions $\mathcal{D}$, $\mathcal{E}$ of a rational function F are *weakly equivalent* if there exists a chain of maximal decompositions $\mathcal{F}_i$, $1 \leqslant i \leqslant s$, of F such that $\mathcal{F}_1 = \mathcal{D}$, $\mathcal{F}_s \sim \mathcal{E}$, and $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$, $1 \leqslant i \leqslant s-1$, by replacing two successive functions $A \circ B$ in $\mathcal{F}_i$ by new functions $C \circ D$ such that $A \circ C = B \circ D$. It is easy to see that this is indeed an equivalence relation. We will denote this equivalence relation by the symbol $\sim_w$. Say that a closed class of rational functions $\mathcal{R}$ is *a Ritt class* if for any $F \in \mathcal{R}$ any two maximal decompositions of F are weakly equivalent. Finally, say that a double decomposition

$$H = A \circ C = B \circ D \tag{43}$$

of a rational function H is *special* if C, D are indecomposable, the pair A, B is reducible, and there exist no rational functions $\tilde{A}$, $\tilde{B}$, U, $\deg U > 1$, such that

$$A = U \circ \tilde{A}, \qquad B = U \circ \tilde{B}, \qquad \tilde{A} \circ C = \tilde{B} \circ D. \tag{44}$$

For decompositions

$$\mathcal{A}\colon\ A = A_r \circ A_{r-1} \circ \cdots \circ A_1, \qquad \mathcal{B}\colon\ B = B_s \circ B_{s-1} \circ \cdots \circ B_1$$

of rational functions A and B denote by $\mathcal{A} \circ \mathcal{B}$ the decomposition

$$A_r \circ A_{r-1} \circ \cdots \circ A_1 \circ B_s \circ B_{s-1} \circ \cdots \circ B_1$$

of the rational function $A \circ B$. In case if a rational function R is indecomposable we will denote the corresponding maximal decomposition by the same letter.

Theorem 5.1. *Let $\mathcal{R}$ be a closed class of rational functions. Suppose that for any $P \in \mathcal{R}$ and any special double decomposition*

$$P = V \circ V_1 = W \circ W_1$$

of P the following condition holds: for any maximal decomposition $\mathcal{V}$ of V and any maximal decomposition $\mathcal{W}$ of W the maximal decompositions $\mathcal{V} \circ V_1$ and $\mathcal{W} \circ W_1$ of P are weakly equivalent. Then $\mathcal{R}$ is a Ritt class.

Proof. For a function $H \in \mathcal{R}$ denote by $d(H)$ the maximal possible length of a maximal decomposition of H. We use the induction on $d(H)$.

If $d(H) = 1$, then any two maximal decompositions of H are weakly equivalent. So, assume that $d(H) > 1$ and let

$$\mathcal{H}_1\colon\ H = F_r \circ F_{r-1} \circ \cdots \circ F_1, \qquad \mathcal{H}_2\colon\ H = G_s \circ G_{s-1} \circ \cdots \circ G_1$$

be two maximal decompositions of a function $H \in \mathcal{R}$. Set

$$F = F_r \circ F_{r-1} \circ \cdots \circ F_2, \qquad G = G_s \circ G_{s-1} \circ \cdots \circ G_2 \tag{45}$$

and consider the double decomposition

$$H = F \circ F_1 = G \circ G_1. \tag{46}$$

If the pair F, G is irreducible, then Theorem 3.6 implies that $\mathcal{H}_1 \sim_w \mathcal{H}_2$ and therefore we must consider only the case when the pair F, G is reducible.

If (46) is special then $\mathcal{H}_1 \sim_w \mathcal{H}_2$ in view of the assumption of the theorem. Thus, assume that (46) is not special and let $\tilde{F}$, $\tilde{G}$, U, $\deg U > 1$, be rational functions such that

$$F = U \circ \tilde{F}, \qquad G = U \circ \tilde{G}, \qquad \tilde{F} \circ F_1 = \tilde{G} \circ G_1.$$

Denote by $\hat{\mathcal{H}}_1, \hat{\mathcal{H}}_2$ the maximal decompositions (45) of the functions F and G and pick some maximal decompositions

$$\tilde{\mathcal{F}}\colon\ \tilde{F} = \tilde{F}_n \circ \tilde{F}_{n-1} \circ \cdots \circ \tilde{F}_1, \qquad \tilde{\mathcal{G}}\colon\ \tilde{G} = \tilde{G}_m \circ \tilde{G}_{\tilde{m}-1} \circ \cdots \circ \tilde{G}_1,$$
$$\mathcal{U}\colon\ U = U_l \circ U_{l-1} \circ \cdots \circ U_1$$

of the functions $\tilde{F}$, $\tilde{G}$, U.

Since $\mathcal{R}$ is closed, $F, G \in \mathcal{R}$. Furthermore, $d(F), d(G) < d(H)$. Therefore, the induction assumption implies that

$$\hat{\mathcal{H}}_1 \sim_w \mathcal{U} \circ \tilde{\mathcal{F}}, \qquad \hat{\mathcal{H}}_2 \sim_w \mathcal{U} \circ \tilde{\mathcal{G}}$$

and hence

$$\mathcal{H}_1 \sim_w \mathcal{U} \circ \tilde{\mathcal{F}} \circ F_1, \qquad \mathcal{H}_2 \sim_w \mathcal{U} \circ \tilde{\mathcal{G}} \circ G_1. \tag{47}$$

Similarly, the function $\tilde{H} = \tilde{F} \circ F_1 = \tilde{G} \circ G_1$ is contained in $\mathcal{R}$ and $d(\tilde{H}) < d(H)$. Hence,

$$\tilde{\mathcal{F}} \circ F_1 \sim_w \tilde{\mathcal{G}} \circ G_1. \tag{48}$$

Now (47) and (48) imply that $\mathcal{H}_1 \sim_w \mathcal{H}_2$. $\square$

As an illustration of our approach let us prove the first Ritt theorem.

Corollary 5.2. *The class $\mathcal{R}_1$ is a Ritt class.*

Proof. In view of Theorem 5.1 it is enough to prove that a polynomial H has no special double decompositions (43). Thus, assume that the pair A, B in (43) is reducible. By Corollary 4.1 there exist polynomials A_1, B_1, U, V such that

$$A = A_1 \circ W, \qquad B = B_1 \circ V, \qquad \deg A_1 = \deg B_1 > 1.$$

Furthermore, Corollary 4.5 implies that

$$A_1 \circ (W \circ C) \sim B_1 \circ (V \circ D).$$

Therefore, equalities (44) hold for

$$U = A_1, \qquad \tilde{A} = W, \qquad \tilde{B} = \mu \circ V,$$

and an appropriate $\mu \in \mathrm{Aut}(\mathbb{CP}^1)$, and hence (43) is not special. $\square$

6. Solutions of Eqs. (9) and (10)

In this section we solve Eqs. (9) and (10).

Lemma 6.1. *Let L_1, L_2 be Laurent polynomials such that the equality*

$$L_1 \circ z^{d_1} = L_2 \circ z^{d_2} \tag{49}$$

holds for some $d_1, d_2 \geqslant 1$. Then there exists a Laurent polynomial R such that

$$L_1 = R \circ z^{D/d_1}, \qquad L_2 = R \circ z^{D/d_2}, \tag{50}$$

where $D = \mathrm{LCM}(d_1, d_2)$.

Proof. For any subgroup G of $\mathrm{Aut}(\mathbb{CP}^1)$ the set k_G, consisting of rational functions f such that $f \circ \sigma = f$ for all $\sigma \in G$, is a subfield k_G of $\mathbb{C}(z)$. Therefore, by the Lüroth theorem k_G has the form $k_G = \mathbb{C}(\varphi_G(z))$ for some rational function φ_G.

Denote by F the Laurent polynomial defined by equality (49). It follows from (49) that F is invariant with respect to the automorphisms $\alpha_1 : z \to \exp(2\pi i/d_1)z$, $\alpha_2 : z \to \exp(2\pi i/d_2)z$. Therefore, F is invariant with respect to the automorphism group G generated by α_1, α_2. Clearly, $\varphi_G = z^D$ and hence $F = R \circ z^D$ for some Laurent polynomial R. Now equalities (50) follow from equalities

$$R \circ z^D = \left(R \circ z^{D/d_1}\right) \circ z^{d_1} = L_1 \circ z^{d_1}, \qquad R \circ z^D = \left(R \circ z^{D/d_2}\right) \circ z^{d_2} = L_2 \circ z^{d_2}. \qquad \square$$

Notice that Lemma 6.1 implies that if A, B, L_1, L_2 is a solution of Eq. (10), then condition 1) of Theorem 1.1 holds.

Set

$$D_n = \frac{1}{2}\left(z^n + \frac{1}{z^n}\right).$$

Notice that for any $m|n$

$$D_n = T_{n/m} \circ D_m = D_{n/m} \circ z^m.$$

Lemma 6.2. *Let F be a rational function such that*

$$F(z) = F(1/z) = F(\varepsilon z),$$

where ε is a root of unity of order $n \geqslant 1$. Then there exists a rational function R such that $F = R \circ D_n$.

Proof. Let G_1 be a subgroup of $\mathrm{Aut}(\mathbb{CP}^1)$ generated by the automorphism $\alpha_1 : z \to \nu z$, where $\nu = \exp(2\pi i/n)$, G_2 be a subgroup of $\mathrm{Aut}(\mathbb{CP}^1)$ generated by the automorphism $\alpha_2 : z \to \frac{1}{z}$, and G_3 be a subgroup of $\mathrm{Aut}(\mathbb{CP}^1)$ generated by α_1 and α_2. It is easy to see that generators of the corresponding invariant fields are $\varphi_{G_1} = z^n$, $\varphi_{G_2} = D_1$, and $\varphi_{G_3} = D_n$. Since F is invariant with respect to G_1 and G_2 it is invariant with respect to G_3 and therefore $F = R \circ D_n$ for some rational function R. $\square$

Lemma 6.3. *Let A, B be polynomials of the same degree and L_1, L_2 be Laurent polynomials such that*

$$A \circ L_1 = B \circ L_2 \tag{51}$$

and $A \circ L_1 \not\sim B \circ L_2$. Then there exist polynomials w_1, w_2 of degree one, a root of unity ν, and $a \in \mathbb{C}$ such that

$$w_1 \circ L_1 \circ (az) = D_r, \qquad w_2 \circ L_2 \circ (az) = D_r \circ (\nu z). \tag{52}$$

Furthermore, if a polynomial A and a Laurent polynomial L satisfy the equation

$$D_r = A \circ L \tag{53}$$

for some $r \geqslant 1$, then there exist a polynomial w of degree one, a root of unity ν, and $n \geqslant 1$ such that

$$w \circ L = D_n \circ (\nu z). \tag{54}$$

Proof. Indeed, it follows from Corollary 4.4 that there exist a rational function W and $\alpha_0, \alpha_1, \alpha_2, \beta_0, \gamma \in \mathbb{C}$ such that

$$L_1 = \left(\alpha_0 + \alpha_1 z + \frac{\alpha_2}{z}\right) \circ W, \qquad L_2 = \left(\beta_0 + \alpha_1 \nu_1 \gamma z + \frac{\alpha_2 \nu_2 \gamma}{z}\right) \circ W,$$

for some mth roots of unity ν_1, ν_2, where $m = \deg A = \deg B$. Furthermore, it follows from $A \circ L_1 \not\sim B \circ L_2$ that $\alpha_1 \alpha_2 \neq 0$. Since the function defined by equality (51) has two poles, this implies that $W = cz^r$, $c \in \mathbb{C}$, and without loss of generality we may assume that $c = 1$. The first part of the lemma follows now from the equalities

$$\alpha_0 + \alpha_1 z^r + \frac{\alpha_2}{z^r} = \left(\alpha_0 + \frac{2\alpha_1 z}{a^r}\right) \circ \frac{1}{2}\left(z^r + \frac{1}{z^r}\right) \circ (az),$$

$$\beta_0 + \alpha_1 \nu_1 \gamma z^r + \frac{\alpha_2 \nu_2 \gamma}{z^r} = \left(\beta_0 + \frac{2\alpha_1 \nu_1 \gamma z}{a^r \nu^r}\right) \circ \frac{1}{2}\left(z^r + \frac{1}{z^r}\right) \circ (\nu a z),$$

where a and ν are numbers satisfying $a^{2r} = \alpha_1/\alpha_2$ and $\nu^{2r} = \nu_1/\nu_2$.

Suppose now that equality (53) holds. Set $n = \deg L_1$ and consider the equality

$$D_r = T_{r/n} \circ D_n = A \circ L. \tag{55}$$

If the decompositions appeared in (55) are not equivalent, then arguing as above and taking into account that in this case $a = 1$, we conclude that (54) holds for some root of unity ν. On the other hand, if the decompositions in (55) are equivalent then (54) holds for $\nu = 1$. $\square$

The theorem below provides a description of solutions of Eq. (9) and implies that if A, L_1, L_2, z^d is a solution of (9) then either condition 1) or condition 4) of Theorem 1.1 holds.

Theorem 6.4. *Suppose that polynomials A, D and Laurent polynomials L_1, L_2 (which are not polynomials) satisfy the equation*

$$A \circ L_1 = L_2 \circ D. \tag{56}$$

Then there exist polynomials R, $\tilde{A}$, $\tilde{D}$, W and Laurent polynomials $\tilde{L}_1$, $\tilde{L}_2$ such that

$$A = R \circ \tilde{A}, \qquad L_2 = R \circ \tilde{L}_2, \qquad L_1 = \tilde{L}_1 \circ W, \qquad D = \tilde{D} \circ W,$$
$$\tilde{A} \circ \tilde{L}_1 = \tilde{L}_2 \circ \tilde{D} \tag{57}$$

and either

$$\tilde{A} \circ \tilde{L}_1 \sim z^n \circ z^r L(z^n), \qquad \tilde{L}_2 \circ \tilde{D} \sim z^r L^n(z) \circ z^n, \tag{58}$$

where L is a Laurent polynomial, $r \geqslant 0$, $n \geqslant 1$, and $\mathrm{GCD}(r,n) = 1$, or

$$\tilde{A} \circ \tilde{L}_1 \sim T_n \circ D_m, \qquad \tilde{L}_2 \circ \tilde{D} \sim D_m \circ z^n, \tag{59}$$

where T_n is the nth Chebyshev polynomial, $n \geqslant 1$, $m \geqslant 1$, and $\mathrm{GCD}(m,n) = 1$.

Proof. Without loss of generality we may assume that $\mathbb{C}(L_1, D) = \mathbb{C}(z)$. Since the function defined by equality (56) has two poles, $D = cz^n$, where $c \in \mathbb{C}$, and we may assume that $c = 1$. Therefore,

$$A \circ L_1 = L_2 \circ D = L_2 \circ D \circ \varepsilon z = A \circ L_1 \circ \varepsilon z,$$

where $\varepsilon = \exp(2\pi i/n)$.

If the decompositions $A \circ L_1$ and $A \circ (L_1 \circ \varepsilon z)$ are equivalent then we have

$$L_1 \circ \varepsilon z = \nu \circ L_1, \tag{60}$$

where $\nu \in \mathrm{Aut}(\mathbb{CP}^1)$. Furthermore, since ν transforms infinity to infinity, ν is a linear function and equality (60) implies that $\nu^{\circ n} = z$. Therefore, $\nu = \alpha + \omega z$ for some nth root of unity ω and $\alpha \in \mathbb{C}$. Now the comparison of the coefficients of both parts of equality (60) implies that L_1 has the form

$$L_1 = \beta + z^r L(z^n), \quad 0 \leqslant r < n,$$

where L is a Laurent polynomial and $\beta \in \mathbb{C}$. Clearly, without loss of generality we may assume that $\beta = 0$ and this implies that also $\alpha = 0$.

It follows from

$$A \circ L_1 = A \circ L_1 \circ \varepsilon z = A \circ \omega z \circ L_1$$

that $A \circ \omega z = A$. Since $\omega = \varepsilon^r$ and $\mathrm{GCD}(r,n) = 1$ in view of the assumption $\mathbb{C}(L_1, D) = \mathbb{C}(z)$, this implies that $A = R \circ z^n$ for some polynomial R. It follows now from the equality

$$L_2 \circ z^n = A \circ L_1 = R \circ z^n \circ z^r L(z^n) = R \circ z^r L^n(z) \circ z^n$$

that $L_2 = R \circ z^r L^n(z)$. Therefore, if the decompositions $A \circ L_1$ and $A \circ (L_1 \circ \varepsilon z)$ are equivalent, then equalities (57), (58) hold.

Suppose now that the decompositions $A \circ L_1$ and $A \circ (L_1 \circ \varepsilon z)$ are not equivalent. Since for any $a \in \mathbb{C}$ we have $z^n \circ (az) = (a^n z) \circ z^n$, it follows from Lemma 6.3 that without loss of generality we may assume that D is still equal z^n while

$$L_1 = D_m = D_1 \circ z^m. \tag{61}$$

Moreover, $\mathrm{GCD}(m,n)=1$ in view of the assumption $\mathbb{C}(L_1, D)=\mathbb{C}(z)$. It follows now from (56) and (61) and Lemmas 6.1 and 6.2 that the Laurent polynomial L defined by equality (56) has the form $L = R \circ D_{nm}$, where R is a polynomial. Therefore,

$$A \circ D_m = R \circ D_{nm} = R \circ T_n \circ D_m$$

and hence $A = R \circ T_n$. Similarly,

$$L_2 \circ z^n = R \circ D_{nm} = R \circ D_m \circ z^n$$

and hence $L_2 = R \circ D_m$. $\square$

7. Reduction of Eq. (8) for reducible pairs A, B

In this section we show that the description of solutions of Eq. (8) for reducible pairs A, B reduces either to the irreducible case or to the description of double decompositions of the function D_n.

Lemma 7.1. *Suppose that polynomials A, B satisfy the equation*

$$A \circ D_n \circ (\mu z) = B \circ D_m, \tag{62}$$

where $\gcd(n,m)=1$ and μ is a root of unity. Then there exist a polynomial R and $l \geqslant 1$ such that $\mu^{2nml}=1$ and

$$A = R \circ \mu^{nml} T_{lm}, \qquad B = R \circ T_{ln}.$$

Proof. Let F be a Laurent polynomial defined by equality (62). It follows from $F = B \circ D_m$ that $F \circ (1/z) = F$. On the other hand,

$$\begin{aligned} F \circ (1/z) &= A \circ D_n \circ (\mu/z) = A \circ \frac{1}{2}\left(\left(\frac{\mu}{z}\right)^n + \left(\frac{z}{\mu}\right)^n\right) \\ &= A \circ D_n \circ (z/\mu) = A \circ D_n \circ (\mu z) \circ \left(z/\mu^2\right) = F \circ \left(z/\mu^2\right). \end{aligned}$$

Therefore, $F = \tilde{F} \circ z^d$ for some rational function $\tilde{F}$ and d equal to the order of $1/\mu^2$. Since also

$$D_n \circ (\mu z) = \frac{1}{2}\left(\mu^n z + \frac{1}{\mu^n z}\right) \circ z^n, \qquad D_m = D_1 \circ z^m,$$

Lemmas 6.1 and 6.2 imply that $F = R \circ D_{nml}$, where R is a rational function and $l = \mathrm{lcm}(d, nm)/nm$.

It follows now from

$$B \circ D_m = R \circ D_{nml} = R \circ T_{ln} \circ D_m$$

that $B = R \circ T_{ln}$. On the other hand, taking into account that $\mu^{nml} = \pm 1$, we have

$$A \circ D_n = F \circ (z/\mu) = R \circ D_{nml} \circ (z/\mu) = R \circ \mu^{nml} D_{nml} = R \circ \mu^{nml} T_{lm} \circ D_n$$

and therefore $A = R \circ (\mu^{nml} T_{lm})$. $\square$

Theorem 7.2. *Suppose that polynomials A, B and Laurent polynomials L_1, L_2 satisfy the equation*

$$A \circ L_1 = B \circ L_2 \tag{63}$$

and the pair A, B is reducible. Then there exist polynomials R, $\tilde{A}$, $\tilde{B}$, W and Laurent polynomials $\tilde{L}_1$, $\tilde{L}_2$ such that

$$A = R \circ \tilde{A}, \qquad B = R \circ \tilde{B}, \qquad L_1 = \tilde{L}_1 \circ W, \qquad L_2 = \tilde{L}_2 \circ W,$$
$$\tilde{A} \circ \tilde{L}_1 = \tilde{B} \circ \tilde{L}_2 \tag{64}$$

and either the pair $\tilde{A}, \tilde{B}$ is irreducible or

$$\tilde{A} \circ \tilde{L}_1 \sim -T_{nl} \circ \frac{1}{2}\left(\varepsilon z^m + \frac{1}{\varepsilon z^m}\right), \qquad \tilde{B} \circ \tilde{L}_2 \sim T_{ml} \circ \frac{1}{2}\left(z^n + \frac{1}{z^n}\right), \tag{65}$$

where T_{nl}, T_{ml} are the corresponding Chebyshev polynomials with $n, m \geqslant 1$, $l > 2$, $\varepsilon^{nl} = -1$, and $\mathrm{GCD}(n, m) = 1$.

Proof. Without loss of generality we may assume that $\mathbb{C}(L_1, L_2) = \mathbb{C}(z)$ and that there exist no rational functions R, $\tilde{A}$, $\tilde{B}$ with $\deg R > 1$ such that the equalities

$$A = R \circ \tilde{A}, \qquad B = R \circ \tilde{B}, \qquad \tilde{A} \circ L_1 = \tilde{B} \circ L_2 \tag{66}$$

hold. If the pair A, B is irreducible, then the statement of the theorem is true, so assume that it is reducible.

By Theorem 3.5 and Corollary 4.1 there exist polynomials A_1, B_1, U, V such that

$$A = A_1 \circ U, \qquad B = B_1 \circ V, \qquad \deg A_1 = \deg B_1 > 1. \tag{67}$$

Furthermore,

$$A_1 \circ (U \circ L_1) \not\sim B_1 \circ (V \circ L_2) \tag{68}$$

since otherwise (66) holds for

$$R = A_1, \qquad \tilde{A} = U, \qquad \tilde{B} = \mu \circ V,$$

where μ is an appropriate automorphism of the sphere. Therefore, by the first part of Lemma 6.3, we may assume without loss of generality that

$$U \circ L_1 = D_r \circ (\nu z), \qquad V \circ L_2 = D_r, \tag{69}$$

where ν is a root of unity. Applying now the second part of Lemma 6.3 to equalities (69) we see that without loss of generality we may assume that

$$L_1 = D_m \circ (\mu z), \qquad L_2 = D_n, \tag{70}$$

where μ is a root of unity. Moreover, $\mathrm{GCD}(n, m) = 1$ in view of the condition $\mathbb{C}(L_1, L_2) = \mathbb{C}(z)$. In particular, we may assume that n is odd.

It follows from (70) by Lemma 7.1 taking into account the assumption about solutions of (66) that there exists a polynomial R of degree one such that

$$A = R \circ \left(\varepsilon^{nl} T_{nl}\right), \qquad L_1 = \frac{1}{2}\left(\varepsilon z^m + \frac{1}{\varepsilon z^m}\right), \qquad B = R \circ T_{ml}, \qquad L_2 = D_n,$$

where $\varepsilon = \mu^m$ and $l \geqslant 1$. Furthermore, since the pair A, B is reducible it follows from Proposition 3.1 that $l > 1$. Clearly, $\varepsilon^{2nl} = 1$. Notice finally that we may assume that $\varepsilon^{nl} = -1$. Indeed, if $\varepsilon^{nl} = 1$ and nl is odd, then, taking into account that $T_{nl} \circ (-z) = -T_{nl}$, we may just change ε to $-\varepsilon$. On the other hand, if nl is even, then $\varepsilon^{nl} = 1$ contradicts to the assumption about solutions of (66). Indeed, since by the assumption n is odd, if nl is even then l is also even and $\varepsilon^{nl} = 1$ implies that $\mu^{mn(l/2)} = \pm 1$. Hence,

$$T_{nl} = T_2 \circ \left(\mu^{mn(l/2)} T_{n(l/2)}\right)$$

and

$$A = (R \circ T_2) \circ \left(\mu^{mn(l/2)} T_{n(l/2)}\right), \qquad B = (R \circ T_2) \circ T_{m(l/2)},$$

where

$$\left(\mu^{mn(l/2)} T_{n(l/2)}\right) \circ D_m \circ (\mu z) = \left(\mu^{mn(l/2)} D_{mn(l/2)}\right) \circ (\mu z) = D_{mn(l/2)} = T_{m(l/2)} \circ D_n.$$

In order to finish the proof we only must show that the algebraic curve

$$T_{ln}(x) + T_{lm}(y) = 0, \tag{71}$$

where $\mathrm{GCD}(n, m) = 1$, is reducible if and only if $l > 2$. First observe that if l is divisible by an odd number f then (71) is reducible since

$$T_{ln}(x) + T_{lm}(y) = T_f \circ T_{n(l/f)} - T_f \circ (-T_{m(l/f)}).$$

Similarly, if l is divisible by 4 then (71) is also reducible since the curve $T_4(x) + T_4(y) = 0$ is reducible.

On the other hand, if $l = 2$ then (71) is irreducible. Indeed, otherwise Corollaries 4.2, 4.1 imply that

$$T_{2n} = A_1 \circ C, \qquad -T_{2m} = B_1 \circ D, \tag{72}$$

for some polynomials A_1, B_1, C, D such that $\deg A_1 = \deg B_1 = 2$ and the curve

$$A_1(x) - B_1(y) = 0 \tag{73}$$

is reducible. Since $T_{2k} = T_2 \circ T_k$ it follows from Corollary 4.5 that if equalities (72) hold then $A_1 = T_2 \circ \mu_1$, $B_1 = -T_2 \circ \mu_2$ for some automorphisms of the sphere. However, it is easy to see that in this case curve (73) is not reducible. Therefore, the condition that equality (72) holds and the condition that curve (73) is reducible may not be satisfied simultaneously and hence (71) is irreducible. □

8. Solutions of Eq. (8) for irreducible pairs A, B

In this section we describe solutions of Eq. (8) in the case when the pair A, B is irreducible. We start from a general description of the approach to the problem.

First of all, if A, B is an irreducible pair of polynomials, then rational functions C, D satisfying Eq. (1) exist if and only if the genus of curve (2) equals zero. Furthermore, it follows from Theorem 2.2 that if $\tilde{C}$, $\tilde{D}$ is a rational solution of (1) such that $\deg \tilde{C} = \deg B$, $\deg \tilde{D} = \deg A$, then for any other rational solution C, D of (1) there exist rational functions C_1, D_1, W such that

$$C = C_1 \circ W, \qquad D = D_1 \circ W, \qquad A \circ C_1 \sim A \circ \tilde{C}, \qquad B \circ D_1 \sim B \circ \tilde{D}.$$

Finally, if C, D are Laurent polynomials, then the function h_1 from Theorem 2.2 should have two poles. On the other hand, it follows from the description of the monodromy of h_1, taking into account that A, B are polynomials, that the number of poles of h_1 equals $\mathrm{GCD}(\deg A, \deg B)$.

The remarks above imply that in order to describe solutions of Eq. (8) for irreducible pairs of polynomials A, B we must describe all irreducible pairs of polynomials A, B such that $\mathrm{GCD}(\deg A, \deg B) \leqslant 2$ and the expression for the genus of (2) provided by formula (21) gives zero. Besides, for each of such pairs we must find a pair of Laurent polynomials $\tilde{L}_1$, $\tilde{L}_2$ satisfying (8) and such that $\deg \tilde{L}_1 = \deg B$, $\deg \tilde{L}_2 = \deg A$.

The final result is the following statement which supplements (over the field $\mathbb{C}$) Theorem 6.1 of the paper of Bilu and Tichy [3].

Theorem 8.1. *Suppose that polynomials A, B and Laurent polynomials L_1, L_2 satisfy the equation*

$$A \circ L_1 = B \circ L_2$$

and the pair A, B is irreducible. Then there exist polynomials $\tilde{A}$, $\tilde{B}$, μ, $\deg \mu = 1$, and rational functions $\tilde{L}_1$, $\tilde{L}_2$, W such that

$$A = \mu \circ \tilde{A}, \qquad B = \mu \circ \tilde{B}, \qquad L_1 = \tilde{L}_1 \circ W, \qquad L_2 = \tilde{L}_2 \circ W, \qquad \tilde{A} \circ \tilde{L}_1 = \tilde{B} \circ \tilde{L}_2$$

and, up to a possible replacement of A to B and L_1 to L_2, one of the following conditions holds:

$$1) \qquad \tilde{A} \circ \tilde{L}_1 \sim z^n \circ z^r R(z^n), \qquad \tilde{B} \circ \tilde{L}_2 \sim z^r R^n(z) \circ z^n,$$

where R is a polynomial, $r \geqslant 0$, $n \geqslant 1$, and $\mathrm{GCD}(n, r) = 1$;

$$2) \qquad \tilde{A} \circ \tilde{L}_1 \sim T_n \circ T_m, \qquad \tilde{B} \circ \tilde{L}_2 \sim T_m \circ T_n,$$

where T_n, T_m are the corresponding Chebyshev polynomials with $m, n \geqslant 1$, and $\mathrm{GCD}(n, m) = 1$;

$$3) \qquad \tilde{A} \circ \tilde{L}_1 \sim -T_{2n_1} \circ \frac{1}{2}\left(\varepsilon z^{m_1} + \frac{1}{\varepsilon z^{m_1}}\right), \qquad \tilde{B} \circ \tilde{L}_2 \sim T_{2m_1} \circ \frac{1}{2}\left(z^{n_1} + \frac{1}{z^{n_1}}\right),$$

where T_{2n_1}, T_{2m_1} are the corresponding Chebyshev polynomials with $m_1, n_1 \geqslant 1$, $\varepsilon^{2n_1} = -1$, and $\mathrm{GCD}(n_1, m_1) = 1$;

$$4) \qquad \tilde{A} \circ \tilde{L}_1 \sim z^2 \circ \frac{z^2 - 1}{z^2 + 1} S\left(\frac{2z}{z^2 + 1}\right), \qquad \tilde{B} \circ \tilde{L}_2 \sim \left(1 - z^2\right) S^2(z) \circ \frac{2z}{z^2 + 1},$$

where S is a polynomial;

$$5) \qquad \tilde{A} \circ \tilde{L}_1 \sim \left(z^2 - 1\right)^3 \circ \frac{3(3z^4 + 4z^3 - 6z^2 + 4z - 1)}{(3z^2 - 1)^2},$$
$$\tilde{B} \circ \tilde{L}_2 \sim \left(3z^4 - 4z^3\right) \circ \frac{4(9z^6 - 9z^4 + 18z^3 - 15z^2 + 6z - 1)}{(3z^2 - 1)^3}.$$

The proof of this theorem is given below and consists of the following stages. First we rewrite formula for the genus of (2) in a more convenient way and prove several related lemmas. Then we introduce the conception of a special value and classify the polynomials having such values. The rest of the proof reduces to the analysis of two cases: the case when one of polynomials A, B does not have special values and the case when both A, B have special values.

Notice that if at least one of polynomials A, B (say A) is of degree 1 then condition 1) holds with $\mu = A,\ R = A^{-1} \circ B,\ n = 1,\ r = 0,\ W = L_2$. So, below we always will assume that $\deg A, \deg B > 1$. Besides, since one can check by a direct calculation that all the pairs of Laurent polynomials $\tilde{L}_1,\ \tilde{L}_2$ in Theorem 8.1 satisfy the requirements above, we will concentrate on the finding of A and B only.

8.1. *Genus formula and related lemmas*

Let $S = \{z_1, z_2, \dots, z_s\}$ be any set of complex numbers which contains all *finite* branch points of a polynomial A of degree n. Then the collection of partitions of the number n:

$$(a_{1,1}, a_{1,2}, \dots, a_{1,p_1}), \dots, (a_{s,1}, a_{s,2}, \dots, a_{s,p_s}),$$

where $(a_{i,1}, a_{i,2}, \dots, a_{i,p_i}),\ 1 \leqslant i \leqslant s$, is the set of lengths of disjoint cycles in the permutation $\alpha_i(A)$, is called the *passport* of A and is denoted by $\mathcal{P}(A)$. Notice that, since we do not require that any of the points of S is a branch point of A, some of partitions above may contain units only. We will call such partitions trivial and will denote by $s(A)$ the number of non-trivial partitions in $\mathcal{P}(A)$.

Below we will assume that S is a union of all finite branch points of a pair of polynomials A, B, $\deg A = n$, $\deg B = m$, and use the notation

$$(b_{1,1}, b_{1,2}, \dots, b_{1,q_1}), \dots, (b_{s,1}, b_{s,2}, \dots, b_{s,q_s}),$$

for the passport $\mathcal{P}(B)$ of B. Clearly, by the Riemann–Hurwitz formula we have

$$\sum_{i=1}^{s} p_i = (s-1)n + 1, \qquad \sum_{i=1}^{s} q_i = (s-1)m + 1. \tag{74}$$

For an irreducible pair of polynomials $A,\ B$ denote by $g(A, B)$ the genus of curve (2). We start from giving a convenient version of formula (21) for $g(A, B)$.

Lemma 8.2.

$$\begin{aligned} -2g(A,B) = {} & \mathrm{GCD}(m,n) - 1 \\ & + \sum_{i=1}^{s} \sum_{j_1=1}^{p_i} \left[a_{i,j_1}(1-q_i) - 1 + \sum_{j_2=1}^{q_i} \mathrm{GCD}(a_{i,j_1} b_{i,j_2}) \right]. \end{aligned} \tag{75}$$

Proof. It follows from (74) that

$$\begin{aligned} \sum_{i=1}^{s} \sum_{j_1=1}^{p_i} \left[a_{i,j_1}(1-q_i) - 1 \right] &= \sum_{i=1}^{s} \left[n(1-q_i) - p_i \right] = ns - n \sum_{i=1}^{s} q_i - \sum_{i=1}^{s} p_i \\ &= ns - n\big((s-1)m + 1\big) - \big((s-1)n + 1\big) = -n(s-1)m - 1. \end{aligned}$$

Therefore, the right side of formula (75) equals

$$-n(s-1)m-2+\sum_{i=1}^{s}\sum_{j_1=1}^{p_i}\sum_{j_2=1}^{q_i}\mathrm{GCD}(a_{i,j_1}b_{i,j_2})+\mathrm{GCD}(m,n).$$

Now (75) follows from (21) taking into account that $r=s+1$. $\square$

Set

$$s_{i,j_1}=a_{i,j_1}(1-q_i)-1+\sum_{j_2=1}^{q_i}\mathrm{GCD}(a_{i,j_1}b_{i,j_2}),$$

$1\leqslant i\leqslant s$, $1\leqslant j_1\leqslant p_i$. Using this notation we may rewrite formula (75) in the form

$$-2g(A,B)=\mathrm{GCD}(m,n)-1+\sum_{i=1}^{s}\sum_{j_1=1}^{p_i}s_{i,j_1}. \tag{76}$$

Two lemmas below provide upper estimates for s_{i,j_1}, $1\leqslant i\leqslant s$, $1\leqslant j_1\leqslant p_i$.

Lemma 8.3. *In the above notation for any fixed pair of indices i, j_1, $1\leqslant i\leqslant s$, $1\leqslant j_1\leqslant p_i$, the following statements hold*:

(a) *If there exist at least three numbers $b_{i,l_1}, b_{i,l_2}, b_{i,l_3}$, $1\leqslant l_1,l_2,l_3\leqslant q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1}\leqslant -2$;*
(b) *If there exist exactly two numbers b_{i,l_1}, b_{i,l_2}, $1\leqslant l_1,l_2\leqslant q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1}\leqslant -1$ and the equality attains if and only if*

$$\mathrm{GCD}(a_{i,j_1}b_{i,l_1})=\mathrm{GCD}(a_{i,j_1}b_{i,l_2})=a_{i,j_1}/2; \tag{77}$$

(c) *If there exists exactly one number b_{i,l_1}, $1\leqslant l_1\leqslant q_i$, which is not divisible by a_{i,j_1} then*

$$s_{i,j_1}=-1+\mathrm{GCD}(a_{i,j_1}b_{i,l_1}). \tag{78}$$

Proof. If there exist at least three numbers $b_{i,l_1}, b_{i,l_2}, b_{i,l_3}$, $1\leqslant l_1,l_2,l_3\leqslant q_i$, which are not divisible by a_{i,j_1} then we have

$$\begin{aligned}s_{i,j_1}&=a_{i,j_1}(1-q_i)-1+\sum_{\substack{j_2=1\\ j_2\neq l_1,l_2,l_3}}^{q_i}\mathrm{GCD}(a_{i,j_1}b_{i,j_2})+\sum_{l_1,l_2,l_3}\mathrm{GCD}(a_{i,j_1}b_{i,l_1})\\ &\leqslant a_{i,j_1}(1-q_i)-1+(q_i-3)a_{i,j_1}+3a_{i,j_1}/2=-a_{i,j_1}/2-1\leqslant -2.\end{aligned}$$

If there exist exactly two numbers b_{i,l_1}, b_{i,l_2}, $1\leqslant l_1,l_2\leqslant q_i$, which are not divisible by a_{i,j_1} then we have

$$\begin{aligned}s_{i,j_1}&=a_{i,j_1}(1-q_i)-1+\sum_{\substack{j_2=1\\ j_2\neq l_1,l_2}}^{q_i}\mathrm{GCD}(a_{i,j_1}b_{i,j_2})+\sum_{l_1,l_2}\mathrm{GCD}(a_{i,j_1}b_{i,l_1})\\ &\leqslant a_{i,j_1}(1-q_i)-1+(q_i-2)a_{i,j_1}+a_{i,j_1}/2+a_{i,j_1}/2=-1,\end{aligned}$$

and the equality attains if and only if

$$\mathrm{GCD}(a_{i,j_1}b_{i,l_1})=\mathrm{GCD}(a_{i,j_1}b_{i,l_2})=a_{i,j_1}/2.$$

Finally, if there exists exactly one number b_{i,l_1} which is not divisible by a_{i,j_1} then we have

$$\begin{aligned} s_{i,j_1} &= a_{i,j_1}(1-q_i) - 1 + \sum_{\substack{j_2=1\\ j_2\neq l_1}}^{q_i} \mathrm{GCD}(a_{i,j_1} b_{i,j_2}) + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}) \\ &= a_{i,j_1}(1-q_i) - 1 + (q_i-1)a_{i,j_1} + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}) = -1 + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}). \qquad \square \end{aligned}$$

Corollary 8.4. *Let B be a polynomial of degree m such that the curve $x^n - B(y) = 0$ is irreducible and of genus zero. Then*

(a) *The equality* $\mathrm{GCD}(n,m) = 1$ *implies that there exists a polynomial ν of degree* 1 *such that $B \circ \nu = z^r R^n$ for some polynomial R and $r \geqslant 1$ such that* $\mathrm{GCD}(r,n) = 1$;
(b) *The equality* $\mathrm{GCD}(n,m) = 2$ *implies that $n = 2$ and there exists a polynomial ν of degree* 1 *such that $B \circ \nu = (1-z^2)S^2$ for some polynomial S.*

Proof. First of all observe that it follows from the irreducibility of $x^n - B(y) = 0$ that among the numbers $b_{1,1}, b_{1,2}, \dots, b_{1,q_1}$ there exists at least one number which is not divisible by n.

If $\mathrm{GCD}(m,n) = 1$, then it follows from formula (76) that $s_{1,1} = 0$, and Lemma 8.3 implies that all the numbers $b_{1,1}, b_{1,2}, \dots, b_{1,q_1}$ but one, say $b_{1,1}$, are divisible by n while $\mathrm{GCD}(n, b_{1,1}) = 1$. Clearly, this implies that $B \circ \nu = z^r R^n$ for some ν, R, and r as above.

Similarly, if $\mathrm{GCD}(m,n) = 2$, then it follows from formula (76) that $s_{1,1} = -1$ and Lemma 8.3 implies that all the numbers $b_{1,1}, b_{1,2}, \dots, b_{1,q_1}$ but two, say $b_{1,1}, b_{1,2}$, are divisible by n while $\mathrm{GCD}(n, b_{1,1}) = \mathrm{GCD}(n, b_{1,2}) = n/2$. Since this implies that $B = z^{n/2} \circ W$ for some polynomial W it follows now from the irreducibility of $x^n - B(y) = 0$ that $n = 2$ and therefore $B \circ \nu = (1-z^2)S^2$ for some ν and S as above. $\square$

Corollary 8.5. *In the notation of Lemma* 8.3 *suppose additionally that*

$$\mathrm{GCD}(b_{i1}, b_{i2}, \dots, b_{iq_i}) = 1. \tag{79}$$

Then the following statements hold:

(a) $s_{i,j_1} \leqslant 0$;
(b) $s_{i,j_1} = 0$ *if and only if either $a_{i,j_1} = 1$ or all the numbers b_{i,j_2}, $1 \leqslant j_2 \leqslant q_i$, except one are divisible by a_{i,j_1}*;
(c) $s_{i,j_1} = -1$ *if and only if $a_{i,j_1} = 2$ and all the numbers b_{i,j_2}, $1 \leqslant j_2 \leqslant q_i$, but two are even.*

Proof. If $a_{i,j_1} = 1$ then $s_{i,j_1} = 0$ so assume that $a_{i,j_1} > 1$. The assumption (79) implies that among the numbers $b_{1,1}, b_{1,2}, \dots, b_{1,q_1}$ there exists at least one number which is not divisible by n. If there exists exactly one number b_{i,l_1} which is not divisible by a_{i,j_1} then in view of (79) necessarily $\mathrm{GCD}(a_{i,j_1} b_{i,l_1}) = 1$ and hence $s_{i,j_1} = 0$ by formula (78). If there exist exactly two numbers b_{i,l_1}, b_{i,l_2}, $1 \leqslant l_1, l_2 \leqslant q_i$, which are not divisible by a_{i,j_1} then it follows from Lemma 8.3 that $s_{i,j_1} \leqslant -1$ where the equality attains if and only if (77) holds. On the other hand, if (77) holds then necessarily $a_{i,j_1} = 2$ since otherwise we obtain a contradiction with (79). Finally, if there exist at least three numbers $b_{i,l_1}, b_{i,l_2}, b_{i,l_3}$, $1 \leqslant l_1, l_2, l_3 \leqslant q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1} \leqslant -2$ by Lemma 8.3. $\square$

8.2. Polynomials with special values

In the notation above say that z_i, $1 \leqslant i \leqslant s$, is a special value of B if

$$\mathrm{GCD}(b_{i,1}, b_{i,2}, \dots, b_{i,q_i}) > 1.$$

It is easy to see that a polynomial P has a special value if and only if there exists $c \in \mathbb{C}$ such that $P - c = z^d \circ R$ for some polynomial R.

Say that z_i, $1 \leqslant i \leqslant s$, is a 1-special value (resp. a 2-special value) of B if all the numbers

$$b_{i,1}, b_{i,2}, \dots, b_{i,q_i}$$

but one (resp. two) are divisible by some number $d > 1$.

Proposition 8.6. *Let B be a polynomial. Then the following statements hold*:

(a) *B may not have two special values, or one special value and one 1-special value, or three 1-special values*;
(b) *If B has two 1-special values then $s(B) = 2$, $\mathcal{P}(B) = \{(1, 2, \dots, 2), (1, 2, \dots, 2)\}$*;
(c) *If B has one 1-special value and one 2-special value then $s(B) = 2$ and either $\mathcal{P}(B) = \{(1, 1, 2), (1, 3)\}$ or $\mathcal{P}(B) = \{(1, 2, 2), (1, 1, 3)\}$.*

Proof. Set $m = \deg B$. Suppose first that B has at least two 1-special values. To be definite assume that these values are z_1, z_2 and that all $(b_{1,1}, \dots, b_{1,q_1})$ but $b_{1,1}$ are divisible by the number d_1 and all $(b_{2,1}, \dots, b_{2,q_2})$ but $b_{2,1}$ are divisible by the number d_2. Then

$$q_1 \leqslant 1 + \frac{m - b_{1,1}}{d_1}, \qquad q_2 \leqslant 1 + \frac{m - b_{2,1}}{d_2}, \tag{80}$$

where the equalities attain if and only if $b_{1,j} = d_1$ for $1 < j \leqslant q_1$ and $b_{2,j} = d_2$ for $1 < j \leqslant q_2$. Furthermore, clearly

$$\sum_{i=1}^{s} q_i \leqslant q_1 + q_2 + (s - 2)m, \tag{81}$$

where the equality attains if and only if $(b_{i,1}, \dots, b_{i,q_i}) = (1, 1, \dots, 1)$ for any $i > 2$. Finally, for $i = 1, 2$ we have

$$q_i \leqslant 1 + \frac{m - b_{i,1}}{d_i} \leqslant 1 + \frac{m - 1}{2} \tag{82}$$

and hence

$$q_1 + q_2 \leqslant 1 + m, \tag{83}$$

where the equality attains only if $d_1 = 2$, $d_2 = 2$, $b_{1,1} = 1$, $b_{2,1} = 1$. Now (81) and (83) imply that

$$\sum_{i=1}^{s} q_i \leqslant (s - 1)m + 1. \tag{84}$$

Since however in view of (74) in this inequality should attain equality we conclude that in all intermediate inequalities should attain equalities. This implies that $s(B) = 2$ and

$$(b_{1,1},\dots,b_{1,q_1})=(1,2,\dots,2),\qquad (b_{2,1},\dots,b_{2,q_1})=(1,2,\dots,2).$$

In particular, we see that B may not have three 1-special values.

In order to prove the first part of the proposition it is enough to observe that if for at least one index 1 or 2, say 1, the corresponding point is special then

$$q_1\leqslant \frac{m}{d_1}\leqslant \frac{m}{2}.$$

Since this inequality is stronger than (82) repeating the argument above we obtain an inequality in (84) in contradiction with (74).

Finally, assume that z_1 is a 1-special value while z_2 is a 2-special value. We will suppose that all $(b_{1,1},\dots,b_{1,q_1})$ but $b_{1,1}$ are divisible by the number d_1 and all $(b_{2,1},\dots,b_{2,q_2})$ but $b_{2,1},b_{2,2}$ are divisible by the number d_2.

If m is odd then $d_2\neq 2$. Hence, in this case $d_2\geqslant 3$,

$$q_1\leqslant 1+\frac{m-b_{1,1}}{d_1}\leqslant 1+\frac{m-1}{2},\qquad q_2\leqslant 2+\frac{m-b_{2,1}-b_{2,2}}{d_2}\leqslant 2+\frac{m-2}{3},$$

and, therefore,

$$q_1+q_2\leqslant \frac{11}{6}+\frac{5m}{6}.$$

If $m>5$ then

$$q_1+q_2\leqslant \frac{11}{6}+\frac{5m}{6}<m+1.$$

Since combined with (81) the last inequality leads to a contradiction with (74) we conclude that $m\leqslant 5$. It follows now from $d_2\geqslant 3$ that necessarily $m=5$ and $(b_{2,1},\dots,b_{2,q_2})=(1,1,3)$. Finally, since z_1 is a 1-special value of B we necessarily have $(b_{1,1},\dots,b_{1,q_1})=(1,2,2)$.

Similarly, if m is even then $d_1\geqslant 3$ and we have

$$q_1\leqslant 1+\frac{m-b_{1,1}}{d_1}\leqslant 1+\frac{m-1}{3},\qquad q_2\leqslant 2+\frac{m-b_{2,1}-b_{2,2}}{d_2}\leqslant 2+\frac{m-2}{2},$$

and

$$q_1+q_2\leqslant \frac{5}{3}+\frac{5m}{6}.$$

If $m>4$ then

$$\frac{5}{3}+\frac{5m}{6}<m+1$$

and as above we obtain a contradiction with (74). On the other hand, if $m\leqslant 4$ then $d_1\geqslant 3$ implies that necessarily $m=4$ and $(b_{1,1},\dots,b_{1,q_1})=(1,3)$. Finally, clearly $(b_{2,1},\dots,b_{2,q_2})=(1,1,2)$. $\square$

8.3. Proof of Theorem 8.1. Part 1

First of all notice that if at least one of polynomials A, B has a unique finite branch point or equivalently is of the form $\mu\circ z^d\circ\nu$ for some $d\geqslant 1$ and polynomials μ,ν of degree one, then it follows from Corollary 8.4 that either condition 1) or condition 4) of Theorem 8.1 holds. So, below we always will assume that both polynomials A, B have at least two finite branch points.

In this subsection we prove Theorem 8.1 under the assumption that at least one of polynomials A, B does not have special values. Without loss of generality we may assume that this polynomial is B. In other words, we may assume that for any i, $1 \leqslant i \leqslant s$, equality (79) holds.

Case 1. Suppose first that $\mathrm{GCD}(n,m)=1$. In this case by formula (76) the condition $g(A,B)=0$ is equivalent to the condition

$$\sum_{i=1}^{s}\sum_{j_1=1}^{p_i} s_{i,j_1}=0. \tag{85}$$

In view of Corollary 8.5(a) this is possible if and only if $s_{i,j_1}=0$, $1 \leqslant i \leqslant s$, $1 \leqslant j_1 \leqslant p_i$.

Since A has at least two finite branch points, Corollary 8.5(a) and Corollary 8.5(b), taking into account that B may not have more than two 1-special values by Proposition 8.6(a), imply that A has exactly two branch points. Furthermore, it follows from Proposition 8.6(b) that $\mathcal{P}(B)$ equals

$$\bigl\{(1,2,2,\dots,2),(1,2,2,\dots,2)\bigr\}. \tag{86}$$

Now Corollary 8.5(b) implies that

$$a_{1,j_1} \leqslant 2, \qquad a_{2,j_2} \leqslant 2, \quad 1 \leqslant j_1 \leqslant p_1,\ 1 \leqslant j_2 \leqslant p_2. \tag{87}$$

Since

$$p_1+p_2=(s-1)n+1=n+1$$

and

$$\sum_{j_1=1}^{p_1} a_{1,j_1}+\sum_{j_1=1}^{p_2} a_{2,j_1}=2n$$

it follows from (87) that among a_{1,j_1}, a_{2,j_2}, $1 \leqslant j_1 \leqslant p_1$, $1 \leqslant j_2 \leqslant p_2$, there are exactly two units and therefore $\mathcal{P}(A)$ equals either (86) or

$$\bigl\{(1,1,2,\dots,2),(2,2,2,\dots,2)\bigr\}. \tag{88}$$

Recall that for any polynomial P such that $\mathcal{P}(P)$ equals (86) or (88) there exist polynomials μ, ν of degree 1 such that $\mu \circ P \circ \nu = T_n$ for some $n \geqslant 1$. A possible way to establish it is to observe that it follows from $T_n(\cos z)=\cos nz$ that T_n satisfies the differential equation

$$n^2\bigl(y^2-1\bigr)=(y')^2\bigl(z^2-1\bigr), \qquad y(1)=1. \tag{89}$$

On the other hand, it is easy to see that if $\mathcal{P}(P)$ equals (86) or (88) and $\deg P=n$ then P satisfies the equation

$$n^2(y-A)(y-B)=(y')^2(z-a)(z-b),$$

for some $A,B,a,b \in \mathbb{C}$ with $y(b)=A$ or B. Therefore for appropriate polynomials μ, ν of degree 1 the polynomial $\mu \circ P \circ \nu$ satisfies Eq. (89) and hence $\mu \circ P \circ \nu = T_n$ by the uniqueness theorem for solutions of differential equations.

Since $\mathcal{P}(B)$ equals (86) and $\mathcal{P}(A)$ equals either (86) or (88) the above characterization of Chebyshev polynomials implies now that if $\mathrm{GCD}(n,m)=1$ then condition 2) holds.

Case 2. If $\mathrm{GCD}(n,m)=2$ then the condition $g(A,B)=0$ is equivalent to the condition that one number from s_{i,j_1}, $1 \leqslant i \leqslant s$, $1 \leqslant j_1 \leqslant p_i$, equals -1 while others equal 0.

Since A has at least two branch points, Corollary 8.5(b) and Corollary 8.5(c), taking into account that if B has two 1-special values, then B does not have 2-special values by Proposition 8.6(b), imply that A has two branch points and B has one 1-special value and one 2-special value. Therefore, by Proposition 8.6(c), $\mathcal{P}(B)$ equals either

$$\{(1,2,2),(1,1,3)\} \tag{90}$$

or

$$\{(1,3),(1,1,2)\}. \tag{91}$$

Furthermore, since the assumption $\mathrm{GCD}(n,m)=2$ implies that $\deg B$ is even we conclude that $\mathcal{P}(B)$ necessarily equals (91). It follows now from Corollary 8.5(b) and Corollary 8.5(c) that for any j_1, $1 \leqslant j_1 \leqslant p_1$, the number a_{1,j_1} equals 1 or 3 and the partition $(a_{2,1},a_{2,2},\dots,a_{2,p_2})$ contains one element equal 2 and others equal 1.

Denote by α (resp. by β) the number of appearances of 1 (resp. of 3) in the first partition of $\mathcal{P}(A)$ and by γ the number of appearances of 1 in the second partition of $\mathcal{P}(A)$. We have

$$\alpha+3\beta=n, \qquad 2+\gamma=n,$$

and, by (74)

$$\alpha+\beta+\gamma=n. \tag{92}$$

The second and the third of the equations above imply that $\alpha+\beta=2$. Hence the partition $(a_{1,1},a_{2,2},\dots,a_{1,p_1})$ is either $(1,3)$ or $(3,3)$ and $\gamma=n-2$ implies that either

$$\mathcal{P}(A)=\mathcal{P}(B)=\{(1,3),(1,1,2)\} \tag{93}$$

or

$$\mathcal{P}(A)=\{(3,3),(2,1,1,1,1)\}. \tag{94}$$

Observe now that for any polynomial R for which $\mathcal{P}(R)$ equals (91) the derivative of R has the form $R'=c(z-a)^2(z-b)$, $a,b,c\in\mathbb{C}$. Therefore, there exist polynomials μ,ν of degree 1 such that

$$\mu\circ R\circ\nu=\int 12z^2(z-1)\,\mathrm{d}z=3z^4-4z^3.$$

Since A and B have the same set of critical values this implies in particular that if (93) holds, then $A=B\circ\lambda$ for some polynomial λ of degree 1 in contradiction with the irreducibility of the curve $A(x)-B(y)=0$. On the other hand, it is easy to see that if equality (94) holds, then there exist polynomials μ,ν_1,ν_2 of degree 1 such that

$$\mu\circ A\circ\nu_1=\left(z^2-1\right)^3, \qquad \mu\circ B\circ\nu_2=3z^4-4z^3.$$

Therefore, if $\mathrm{GCD}(n,m)=2$, then condition 5) holds.

8.4. Proof of Theorem 8.1. Part 2

Suppose now that both polynomials A and B have special values. Then by Proposition 8.6(b) each of them has a unique special value. The special values of A and B either coincide or are different. If they are different then

$$A = \left(z^{d_1} + \beta_1\right) \circ \hat{A}, \qquad B = \left(z^{d_2} + \beta_2\right) \circ \hat{B}, \tag{95}$$

for some $\beta_1, \beta_2 \in \mathbb{C}$, $\beta_1 \neq \beta_2$, and $d_1, d_2 > 1$. Since the pair A, B is irreducible and $g(A, B) = 0$ the pair $A_0 = z^{d_1} + \beta_1$, $B_0 = z^{d_2} + \beta_2$ is also irreducible and

$$g(A_0, B_0) = 0. \tag{96}$$

Formula (75) implies that

$$-2g(A_0, B_0) = d_1 + d_2 - d_1 d_2 + \mathrm{GCD}(d_1, d_2) - 2. \tag{97}$$

If $\mathrm{GCD}(d_1, d_2) = 1$, then (96) is equivalent to the equality $(d_1 - 1)(1 - d_2) = 0$ which is impossible. On the other hand, if $\mathrm{GCD}(d_1, d_2) = 2$, then (96) is equivalent to the equality $(d_1 - 1)(1 - d_2) = -1$ which holds if and only if $d_1 = d_2 = 2$.

Repeatedly using Theorem 3.5 and Corollary 4.1 we can find polynomials P, Q, U, V such that

$$\hat{A} = P \circ U, \qquad \hat{B} = Q \circ V, \qquad \deg P = \deg Q,$$

and the pair U, V is irreducible. Setting

$$A_1 = A_0 \circ P, \qquad B_1 = B_0 \circ Q \tag{98}$$

we see that equality (67) holds. Furthermore, equivalence (68) is impossible since otherwise $A_1 = B_1 \circ \mu$ for some polynomial μ of degree 1 and it follows from Corollary 4.5 and equalities (98) that $A_0 = B_0 \circ \nu$ for some polynomial ν of degree 1 in contradiction with the irreducibility of the pair A_0, B_0. Now using the same reasoning as in the proof of Theorem 7.2 and taking into account that the pair A_0, B_0 is irreducible we conclude that condition 3) holds.

In the case when the special values of A and B coincide we can assume without loss of generality that

$$A = z^{d_1} \circ U, \qquad B = z^{d_2} \circ V, \tag{99}$$

where

$$d_1 = \mathrm{GCD}(a_{1,1}, a_{1,2}, \ldots, a_{1,p_1}) > 1, \qquad d_2 = \mathrm{GCD}(b_{1,1}, b_{1,2}, \ldots, b_{1,q_1}) > 1,$$

and

$$\mathrm{GCD}(d_1, d_2) = 1 \tag{100}$$

in view of the irreducibility of the pair A and B. Notice that, since A and B have at least two critical values, the inequalities $p_1 \geqslant 2$, $q_1 \geqslant 2$ hold. Finally, without loss of generality we may assume that $m = \deg B$ is greater than $n = \deg A$. We will consider the cases $\mathrm{GCD}(d_1, m) = 2$ and $\mathrm{GCD}(d_1, m) = 1$ separately and will show that in both cases there exist no irreducible pairs A, B with $g(A, B) = 0$.

Case 1. Suppose first that $\mathrm{GCD}(d_1, m) = 2$. Then necessarily $\mathrm{GCD}(n, m) = 2$ and, since

$$x^{d_1} - B(y) = 0 \tag{101}$$

is an irreducible curve of genus zero, Corollary 8.4 implies that $d_1 = 2$ and all the numbers $b_{1,1}, b_{1,2}, \dots, b_{1,q_1}$ but two, say b_{1,q_1-1}, b_{1,q_1}, are even while b_{1,q_1-1}, b_{1,q_1} are odd. Since by the assumption each a_{1,j_1}, $1 \leqslant j_1 \leqslant p_1$, is divisible by $d_1 = 2$, this implies in particular that for each j_1, $1 \leqslant j_1 \leqslant p_1$,

$$\mathrm{GCD}(a_{1,j_1} b_{1,q_1-1}) \leqslant a_{1,j_1}/2, \qquad \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \leqslant a_{1,j_1}/2.$$

Returning now to polynomials A, B we conclude that for each j_1, $1 \leqslant j_1 \leqslant p_1$,

$$\begin{aligned} s_{1,j_1} &= a_{1,j_1}(1-q_1) - 1 + \sum_{j_2=1}^{q_1-2} \mathrm{GCD}(a_{1,j_1} b_{1,j_2}) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1-1}) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \\ &\leqslant a_{1,j_1}(1-q_1) - 1 + a_{1,j_1}(q_1-2) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1-1}) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \\ &\leqslant -a_{1,j_1} - 1 + a_{1,j_1}/2 + a_{1,j_1}/2 \leqslant -1. \end{aligned}$$

Since $p_1 \geqslant 2$ and by Corollary 8.5(a) for any i, $1 < i \leqslant s$, and j, $1 \leqslant j_1 \leqslant p_i$, the inequality $s_{i,j_1} \leqslant 0$ holds it follows now from formula (76) that $g(A, B) < 0$.

Case 2. Similarly, if $\mathrm{GCD}(d_1, m) = 1$ then Lemma 8.4 applied to curve (101) implies that each b_{1,j_1}, $1 \leqslant j_1 \leqslant q_1$, except one, say b_{1,q_1}, is divisible by d_1 while $\mathrm{GCD}(b_{1,q_1}, d_1) = 1$ and returning to A, B and taking into account that each a_{1,j_1}, $1 \leqslant j_1 \leqslant p_1$, is divisible by d_1 we obtain that

$$\begin{aligned} s_{1,j_1} &= a_{1,j_1}(1-q_1) - 1 + \sum_{j_2=1}^{q_1-1} \mathrm{GCD}(a_{1,j_1} b_{1,j_2}) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \\ &\leqslant -1 + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \leqslant -1 + a_{1,j_1}/d_1. \end{aligned} \tag{102}$$

Hence,

$$\sum_{j_1=1}^{p_1} s_{1,j_1} \leqslant -p_1 + n/d_1. \tag{103}$$

Furthermore, since each b_{1,j_2}, $1 \leqslant j_2 \leqslant q_1$, except one is divisible by d_1, each b_{1,j_2}, $1 \leqslant j_2 \leqslant q_1$, is divisible by d_2, and equality (100) holds we have

$$(q_1 - 1)d_1 d_2 + d_2 \leqslant m$$

and therefore

$$q_1 \leqslant 1 + m/d_1 d_2 - 1/d_1.$$

Since by (74) the inequality

$$q_1 + q_i \geqslant m + 1 \tag{104}$$

holds for any i, $2 \leqslant i \leqslant s$, this implies that

$$q_i \geqslant m - m/d_1 d_2 + 1/d_1. \tag{105}$$

Denote by γ_i, $2 \leqslant i \leqslant s$, the number of units among the numbers b_{i,j_2}, $1 \leqslant j_2 \leqslant q_i$. Since the number of non-units is $\leqslant m/2$ the inequality $\gamma_i \geqslant q_i - m/2$ holds and therefore (105) implies that

$$\gamma_i \geqslant m/2 - m/d_1d_2 + 1/d_1. \tag{106}$$

For any i, j_1, $2 \leqslant i \leqslant s$, $1 \leqslant j_1 \leqslant p_i$, we have

$$s_{i,j_1} \leqslant a_{i,j_1}(1-q_i) - 1 + a_{i,j_1}(q_i - \gamma_i) + \gamma_i = (1-\gamma_i)(a_{i,j_1} - 1). \tag{107}$$

Since this implies that

$$\sum_{j_1=1}^{p_i} s_{i,j_1} = (1-\gamma_i)\sum_{j_1=1}^{p_i}(a_{i,j_1} - 1) \leqslant (1-\gamma_i)(n-p_i) \tag{108}$$

it follows now from (106) that

$$\sum_{j_1=1}^{p_i} s_{i,j_1} \leqslant \bigl(1 - 1/d_1 + m(1/d_1d_2 - 1/2)\bigr)(n-p_i).$$

Therefore, using (74) we obtain that

$$\sum_{i=2}^{s}\sum_{j_1=1}^{p_i} s_{i,j_1} \leqslant \bigl(1 - 1/d_1 + m(1/d_1d_2 - 1/2)\bigr)(p_1 - 1). \tag{109}$$

Set

$$S = \sum_{i=1}^{s}\sum_{j_1=1}^{p_1} s_{i,j_1}.$$

Since $\mathrm{GCD}(n,m) = 1$ or 2 it follows from formula (76) that in order to finish the proof it is enough to show that $S < -1$.

Since $p_1 \geqslant 2$ it follows from (103), (109) that

$$\begin{aligned} S &\leqslant -p_1 + n/d_1 + \bigl(1 - 1/d_1 + m(1/d_1d_2 - 1/2)\bigr)(p_1 - 1) \\ &= -1 + n/d_1 - \frac{p_1 - 1}{d_1} + m(1/d_1d_2 - 1/2)(p_1 - 1) \\ &< -1 + n/d_1 + m(1/d_1d_2 - 1/2)(p_1 - 1). \end{aligned} \tag{110}$$

If $p_1 \geqslant 3$ then it follows from (110), taking into account the assumption $m \geqslant n$ and the inequality $1/d_1d_2 - 1/2 < 0$, that

$$S < -1 + n(1/d_1 + 2/d_1d_2 - 1).$$

Since $1/d_1 + 2/d_1d_2 - 1 \leqslant 0$ for any $d_1, d_2 \geqslant 2$, this implies that $S < -1$.

If $p_1 = 2$ then (110) implies that

$$S < -1 + n(1/d_1 + 1/d_1d_2 - 1/2).$$

Since $1/d_1 + 1/d_1d_2 - 1/2 \leqslant 0$ whenever $d_1 > 2$ we obtain that $S < -1$ also if $p_1 = 2$ but $d_1 > 2$.

Finally, if $p_1 = 2$, $d_1 = 2$ but $m \geqslant (3/2)n$, then it follows from equality (110) that

$$S < -1 + n(3/4d_2 - 1/4).$$

Since $d_1 = 2$ implies $d_2 \geqslant 3$ in view of (100), we conclude again that $S < -1$.

Therefore, the only case when the proof of the inequality $S < -1$ is still not finished is the one when $p_1 = 2$, $d_1 = 2$, and $n \leqslant m < (3/2)n$. In this case apply the reasoning above to A and B switched keeping the same notation. In other words, assume that $q_1 = 2$, $d_2 = 2$, and

$$2n/3 < m \leqslant n. \tag{111}$$

Then by (104) we have $q_i \geqslant m - 1$ for any i, $2 \leqslant i \leqslant s$. Therefore, corresponding partitions of m are either trivial or have the form $(1, 1, \dots, 1, 2)$ and hence

$$\gamma_i \geqslant m - 2, \quad 2 \leqslant i \leqslant s. \tag{112}$$

It follows now from (108), (112), (74), and (111) that

$$\sum_{i=2}^{s} \sum_{j_1=1}^{p_i} s_{i,j_1} \leqslant (3 - m)(p_1 - 1) < (3 - 2n/3)(p_1 - 1) \leqslant 3 - 2n/3. \tag{113}$$

Since $d_2 = 2$ implies $d_1 \geqslant 3$ in view of (100), it follows now from (113) and $p_1 \geqslant 2$ that

$$S < -p_1 + n/d_1 + 3 - 2n/3 \leqslant 1 + n/d_1 - 2n/3 \leqslant 1 - n/3.$$

If $n \geqslant 6$ then this inequality implies that $S < -1$. On the other hand, the inequality $n \leqslant 5$ is impossible since otherwise equalities (99), (111), and $p_1 \geqslant 2$ imply that $d_1 = d_2 = 2$ in contradiction with (100).

In order to finish the proof of Theorem 8.1 it is enough to notice that for any choice of $\tilde{A}$, $\tilde{B}$ in conditions 1)–5) the curve

$$\tilde{A}(x) - \tilde{B}(y) = 0 \tag{114}$$

is indeed irreducible. For cases 1) and 2) this is a corollary of Proposition 3.1. For case 3) this was proved in the end of the proof of Theorem 7.2. In case 4) corresponding curve (114) is irreducible since otherwise Corollary 4.2 would imply that there exists a polynomial T such that $\tilde{B} = z^2 \circ T$ in contradiction with $\tilde{B} = (1 - z^2)S^2$. Finally, since $\tilde{B}$ in 5) is indecomposable it follows from Corollary 4.2 taking into account Corollary 4.1 that corresponding curve (114) is irreducible.

9. Proof of Theorem 1.1

Since the description of double decompositions of functions from $\mathcal{R}_2$ reduces to the corresponding problem for Laurent polynomial and any double decomposition of a Laurent polynomial is equivalent to (8), (9), or (10), the first part of Theorem 1.1 follows from Theorem 6.4, Theorem 7.2, Theorem 8.1 and Lemma 6.1. The proof of the second part is given below.

Theorem 9.1. *The class $\mathcal{R}_2$ is a Ritt class.*

Proof. We will use Theorem 5.1 and the first part of Theorem 1.1. First observe that the first part of Theorem 1.1 implies that if $A \circ C = B \circ D$ is a double decomposition of a function from $\mathcal{R}_2$ such that C, D are indecomposable and there exist no rational functions $\tilde{A}$, $\tilde{B}$, U, $\deg U > 1$, such that (44) holds then there exist automorphisms of the sphere μ, W and rational functions $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, $\tilde{D}$ such that one of conclusions of Theorem 1.1 holds. Moreover, it was shown above that in cases 1)–3) and 6) the pair $\tilde{A}$, $\tilde{B}$ is irreducible.

Observe now that in case 4) the pair $\tilde{A}$, $\tilde{B}$ is also irreducible. Indeed, since $\mathrm{GCD}(n, m) = 1$, it follows from the construction given in Section 2.2 that for the pair $f = \tilde{A}$, $g = \tilde{B}$ the permutation δ_i, $1 \leqslant i \leqslant r$, corresponding to the loop around the infinity contains two cycles. Therefore, if the pair $\tilde{A}$, $\tilde{B}$ is reducible, then $o(f, g) = 2$ and both functions h_1, h_2 from Theorem 2.2 have a unique pole. On the other hand, the last statement contradicts to the fact that $h_1 = \tilde{B} \circ v_1$, $h_2 = \tilde{B} \circ v_2$ for some rational functions v_1, v_2 since $\tilde{B}$ has two poles.

Finally, as it was observed in the end of the proof of Theorem 7.2, in case 5) the pair $\tilde{A}$, $\tilde{B}$ is reducible whenever $l > 2$. Since in this case $\tilde{C}$ and $\tilde{D}$ are decomposable unless $n = 1$, $m = 1$, it follows now from Theorem 5.1 that in order to prove the proposition it is enough to check that for any choice of maximal decompositions

$$-T_l = u_d \circ u_{d-1} \circ \cdots \circ u_1, \qquad T_l = v_l \circ v_{l-1} \circ \cdots \circ v_1,$$

the decompositions

$$u_d \circ u_{d-1} \circ \cdots \circ u_1 \circ \frac{1}{2}\left(\varepsilon z + \frac{1}{\varepsilon z}\right), \qquad v_l \circ v_{l-1} \circ \cdots \circ v_1 \circ \frac{1}{2}\left(z + \frac{1}{z}\right), \tag{115}$$

where $\varepsilon^l = -1$, are weakly equivalent.

Since $T_l = T_d \circ T_{l/d}$ for any $d|l$, it follows from Corollary 4.5 that any maximal decomposition of T_l is equivalent to $T_l = T_{d_1} \circ T_{d_2} \circ \cdots \circ T_{d_s}$, where $d_1, d_2, \ldots, d_s$ are prime divisors of l such that $d_1 d_2 \cdots d_s = l$. Taking into account that for $d \geqslant 1$

$$T_d \circ \frac{1}{2}\left(z + \frac{1}{z}\right) = \frac{1}{2}\left(z + \frac{1}{z}\right) \circ z^d,$$

this implies easily that both decompositions (115) are weakly equivalent to some decomposition of the form

$$\frac{1}{2}\left(z + \frac{1}{z}\right) \circ z^{d_1} \circ z^{d_2} \circ \cdots \circ z^{d_s}. \qquad \square$$

Acknowledgements

The results of this paper were obtained mostly during the visits of the author to the Max-Planck-Institut für Mathematik in Summer 2005 and Spring 2007 and were partially announced in [19]. Seizing an opportunity the author would like to thank the Max-Planck-Institut for the hospitality. Besides, the author is grateful to Y. Bilu, M. Muzychuk, and M. Zieve for discussions of ideas and preliminary results of this paper before its publication.

References

[1] R. Avanzi, U. Zannier, The equation $f(X) = f(Y)$ in rational functions $X = X(t)$, $Y = Y(t)$, Compos. Math. 139 (3) (2003) 263–295.
[2] Y. Bilu, Quadratic factors of $f(x) - g(y)$, Acta Arith. 90 (4) (1999) 341–355.
[3] Y. Bilu, R. Tichy, The Diophantine equation $f(x) = g(y)$, Acta Arith. 95 (3) (2000) 261–288.
[4] M. Briskin, N. Roytvarf, Y. Yomdin, Center conditions at infinity for Abel differential equation, Ann of Math., in press.

[5] J.M. Couveignes, L. Granboulan, Dessins from a geometric point of view, in: The Grothendieck Theory of Dessins d'Enfants, Cambridge Univ. Press, 1994, pp. 79–113.
[6] H. Engstrom, Polynomial substitutions, Amer. J. Math. 63 (1941) 249–255.
[7] A. Eremenko, Some functional equations connected with the iteration of rational functions, Leningrad Math. J. 1 (4) (1990) 905–919.
[8] J. Gutierrez, D. Sevilla, Building counterexamples to generalizations for rational functions of Ritt's decomposition theorem, J. Algebra 303 (2) (2006) 655–667.
[9] M. Fried, On a theorem of Ritt and related diophantine problems, J. Reine Angew. Math. 264 (1973) 40–55.
[10] M. Fried, Fields of definition of function fields and a problem in the reducibility of polynomials in two variables, Illinois J. Math. 17 (1973) 128–146.
[11] M. Fried, Arithmetical properties of function fields. II. The generalized Schur problem, Acta Arith. 25 (1974) 225–258.
[12] F. Klein, Lectures on the Icosahedron and the Solution of Equations of the Fifth Degree, Dover Publications, New York, 1956.
[13] A. Kurosch, The Theory of Groups, vol. I, Chelsea Publishing Company, New York, 1955.
[14] N. Magot, A. Zvonkin, Belyi functions for Archimedean solids, Discrete Math. 217 (1–3) (2000) 249–271.
[15] W. Massey, Algebraic Topology: An Introduction, Grad. Texts in Math., vol. 56, 1981.
[16] R. Miranda, Algebraic Curves and Riemann Surfaces, Grad. Stud. Math., vol. 5, Amer. Math. Soc., Providence, RI, 1995.
[17] F. Pakovich, On the functional equation $F(A) = G(B)$, where A, B are polynomial and F, G are continuous functions, Math. Proc. Cambridge Philos. Soc. 143 (2) (2007) 469–472.
[18] F. Pakovich, On polynomials sharing preimages of compact sets and related questions, Geom. Funct. Anal. 18 (1) (2008) 163–183.
[19] F. Pakovich, On analogues of Ritt theorems for rational functions with at most two poles, Russian Math. Surveys 63 (2) (2008) 181–182.
[20] F. Pakovich, The algebraic curve $P(x) - Q(y) = 0$ and functional equations, Complex Var. Elliptic Equ., in press, arXiv:0804.0736v2.
[21] F. Pakovich, M. Muzychuk, Solution of the polynomial moment problem, Proc. Lond. Math. Soc. 2009, doi:10.1112/plms/pdp010, in press.
[22] F. Pakovich, On the equation $P(f) = Q(g)$, where P, Q are polynomials and f, g are entire functions, preprint, arXiv:0804.0739.
[23] J. Ritt, Prime and composite polynomials, Trans. Amer. Math. Soc. 23 (1922) 51–66.
[24] J. Ritt, Equivalent rational substitutions, Trans. Amer. Math. Soc. 26 (1924) 221–229.
[25] J. Ritt, Permutable rational functions, Trans. Amer. Math. Soc. 25 (1923) 399–448.
[26] A. Schinzel, Polynomials with Special Regard to Reducibility, Encyclopedia Math. Appl., vol. 77, Cambridge Univ. Press, 2000.
[27] P. Tortrat, Sur la composition des polynômes, Colloq. Math. 55 (2) (1988) 329–353.
[28] U. Zannier, Ritt's second theorem in arbitrary characteristic, J. Reine Angew. Math. 445 (1993) 175–203.