PRIME AND COMPOSITE LAURENT POLYNOMIALS

F. PAKOVICH

Abstract. In his paper [15] Ritt constructed a decomposition theory of polynomials and described explicitly polynomial solutions of the functional equation $f(p(z)) = g(q(z))$. In this paper we construct a self-contained decomposition theory of rational functions with at most two poles. In particular, we give new proofs of the theorems of Ritt and of the theorem of Bilu and Tichy. Besides, we study general properties of the equation above in the case when f, g, p, q are holomorphic functions on compact Riemann surfaces.

1. Introduction

Let $F(z)$ be a rational function with complex coefficients. The function $F(z)$ is called indecomposable if the equality $F = F_1 \circ F_2$, where $F_1(z)$, $F_2(z)$ are rational functions and $F_1 \circ F_2$ denotes a superposition $F_1(F_2(z))$, implies that at least one of the functions $F_1(z), F_2(z)$ is of degree one. Clearly, any rational function $F(z)$ can be decomposed into a composition $F = F_r \circ F_{r-1} \circ \cdots \circ F_1$ of indecomposable rational functions. We will call such decompositions maximal.

In general, a rational function may have many maximal decompositions and the ultimate goal of the decomposition theory of rational functions is to describe a general structure of all such decompositions up to an equivalence, where by definition two decompositions

$$F = F_1 \circ F_2 \circ \cdots \circ F_r \quad \text{and} \quad F = G_1 \circ G_2 \circ \cdots \circ G_r,$$

which may or may not be maximal, are called equivalent if there exist Möbius transformations μ_i, $1 \leq i \leq r-1$, such that

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

Essentially, the unique case when this problem is completely solved is the one investigated by Ritt in his classical paper [15] concerning the situation when $F(z)$ is a polynomial.

The Ritt results can be summarized as a union of two theorems usually called the first and the second Ritt theorems (see [15] and also [18], [17] for the case when the ground field is distinct form $\mathbb{C}$). The first Ritt theorem states that any two maximal decompositions $\mathcal{D}, \mathcal{E}$ of a polynomial $P(z)$ have an equal number of terms and there exists a chain of decompositions $\mathcal{F}_i$, $1 \leq i \leq s$, of $P(z)$ such that $\mathcal{F}_1 = \mathcal{D}$, $\mathcal{F}_s \sim \mathcal{E}$, and $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$, $1 \leq i \leq s-1$, by a change of a segment of $\mathcal{F}_i$ consisting of two consecutive terms $A \circ B$ to a new segment $C \circ D$ such that

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

1991 *Mathematics Subject Classification.* Primary 30D05; Secondary 14H30.

Key words and phrases. Ritt's theorems, decompositions of rational functions.

The first Ritt theorem reduces the decomposition problem of polynomials to a description of indecomposable polynomial solutions of equation (1). The second Ritt theorem gives such a description and states that if (1) holds and the decompositions $A\circ B$ and $C\circ D$ are not equivalent then there exist Möbius transformations $\mu_1(z), \mu_2(z)$ such that

$$A=\mu_1\circ\tilde A,\quad B=\mu_1\circ\tilde B,\quad C=\tilde C\circ\mu_2,\quad D=\tilde D\circ\mu_2$$

and either

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

for the Chebyshev polynomials $T_m(z), T_n(z)$ and $\mathrm{GCD}(n,m)=1$, or

$$\tilde A\circ\tilde B\sim z^n\circ z^mR(z^n),\qquad \tilde C\circ\tilde D\sim z^mR^n(z)\circ z^m,$$

for a polynomial $R(z)$ and $\mathrm{GCD}(n,m)=1$.

For arbitrary rational functions the first Ritt theorem fails to be true. Furthermore, there exist rational functions with maximal decompositions of different lengths. The simplest examples of this phenomenon can be constructed with the use of rational functions $f(z)$ which are the Galois coverings. Notice that all the functions with this property were described by F. Klein in his famous book [8]. They 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 the Platonic solids [4], [10].

The reason for the choice of these functions as possible counterexamples to the first Ritt theorem is the fact that for such a function $f(z)$ maximal decompositions of $f(z)$ correspond to maximal chains of subgroups

$$e=G_r\subset G_{r-1}\subset\ \dots\ \subset G_0=G,$$

where G is the monodromy group of $f(z)$. Therefore, in order to find maximal decompositions of different lengths of f it is enough to find the corresponding chains of subgroups of G. For C_n and D_n such chains do not exist but already for $G=A_4$ they do. The corresponidng decompositions of different lenght of a function $f(z)$ were found explicitely in [5]. If $f(z)$ is normalized to be the Belyi function for the tetrahedron (see [10])

$$f(z)=-64\frac{(z^3+1)^3}{(z^3-8)^3z^3},$$

then these decompositions take especially simple form:

$$f(z)=-64\frac{(z+1)^3}{z(z-8)^3}\circ z^3,\qquad f(z)=-64z^3\circ\frac{z-1}{z^2-4}\circ\frac{z^2+2}{z+1}$$

(M. Zieve communicated to us [19] that actually these examples essentially were mentioned already by Ritt in his paper [16] although Ritt did not write the corresponding decompositions in an explicit form).

Notice that the problem of description of polynomial solutions of (1) is essentially equivalent to the problem of description of the algebraic curves of the form

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

which have a factor of genus zero with one point at infinity. A more general question of description of curves (2) having a factor of genus 0 with at most *two* points at infinity is closely connected to the number theory and was studied in the papers of Fried [6] and Bilu and Tichy [2]. In particular, in [2] an explicit list of such curves was obtained. Another important result concerning functional equation (1), obtained by Avanzi and Zannier [1], gives a description of rational solutions of (1) under condition that $A(z)$ and $B(z)$ are polynomials equal between themselves. Finally, notice that the problem of description of rational solutions of (1) under condition that $C(z)$, $D(z)$ are polynomials is quite simple and essentially reduces to the Ritt theorem [14].

It turns out that a fruitful way to investigate general properties of equation (1) is to study the structure of its possible solutions $C(z), D(z)$ for fixed $A(z), B(z)$. In the first part of this paper we develop this approach in a more general context of holomorphic functions on compact Riemann surfaces. Namely, we investigate the equation

$$h = f \circ p = g \circ q, \tag{3}$$

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 unknown Riemann surface C. In subsection 2.1 we give a description of the general structure of solutions of (3). We show (Theorem 2.1) that there exists a finite number $o(f, g)$ of solutions $h_j(z), p_j(z), q_j(z)$ of (3) such that any other solution has the form

$$h = h_j \circ w, \quad p = \tilde{p} \circ w, \quad q = \tilde{q} \circ w,$$

where $w(z)$ is a holomorphic function and

$$f \circ \tilde{p} \sim f \circ p_j, \qquad g \circ \tilde{q} \sim g \circ q_j.$$

Moreover, we describe explicitly the monodromy of $h_j(z)$ via the monodromy of $f(z), g(z)$.

Theorem 1 naturally distinguishes a class of pairs of holomorphic functions such that $o(f, g) = 1$. We will call the pairs from this class irreducible, since if $f(z)$, $g(z)$ are polynomials then the condition $o(f, g) = 1$ is equivalent to the condition that the algebraic curve $f(x) - g(y) = 0$ is irreducible. In subsection 2.2 we give (Theorem 2.2) a topological criterion for a pair $f(z), g(z)$ to be irreducible. As a corollary we obtain the following result (Theorem 2.3) which generalizes the corresponding result of Fried [7] about polynomials: if a pair of holomorphic functions $f(z), g(z)$ is reducible then there exist holomorphic functions $\tilde{f}(z)$ $\tilde{g}(z)$, $p(z)$, $q(z)$ such that

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

and the normalizations of $\tilde{f}(z)$ and $\tilde{g}(z)$ coincide. We also show (Theorem 2.4) that if (3) is a double decomposition with indecomposable $p(z)$, $q(z)$ and the pair $f(z)$, $g(z)$ is irreducible then $f(z)$, $g(z)$ are indecomposable.

Further, in subsection 2.3 of the first part of the paper we establish (Proposition 2.4) an important property of equation (3) in the case when $f(z), g(z)$ are "generalized" polynomials that is holomorphic functions for which the preimage of infinity contains a unique point. In particular, Proposition 2.4 implies (Corollary 2.5) that if $A(z), B(z)$ are "usual" polynomials of the same degree and $C(z), D(z)$ are rational functions such that equality (1) holds then there exist rational functions $\tilde{C}(z)$, $\tilde{D}(z)$, $W(z)$ such that

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

and $\tilde{C}(z)$ and $\tilde{D}(z)$ have an equal number of poles all of which are simple. This is a generalization of a well known fact that two decompositions $A \circ C$ and $B \circ D$ of a polynomial $P(z)$ for which $\deg A(z) = \deg B(z)$ are equivalent.

Finally, in subsection 2.4 we introduce a notion of a closed class of rational function as of a subset $\mathcal{H}$ of $\mathbb{C}(z)$ such that the condition $G \circ H \in \mathcal{H}$ implies that $G \in \mathcal{H}$, $H \in \mathcal{H}$. For example, for fixed $k \geq 1$ the set $\mathcal{R}_k$ consisting of the rational functions $F(z)$ for which

$$\min_{z \in \mathbb{CP}^1} \sharp\{F^{-1}\{z\}\} \leq k$$

is a closed class and the Ritt theorems can be interpreted as a decomposition theory for the class $\mathcal{R}_1$. We show (Theorem 2.5) that in order to check that the first Ritt theorem holds for all functions from a closed class $\mathcal{H}$ it is enough to check that it holds for the functions from a certain subset of $\mathcal{H}$ related to reducible pairs from $\mathcal{H}$. This criterion is useful since the corresponding subset is considerably less than $\mathcal{H}$. For example, for the class $\mathcal{R}_1$ this subset turns out to be empty that implies in particular the truth of the first Ritt for this class (Proposition 2.5).

In connection with the first Ritt theorem let us mention also the following observation which is a direct corollary of Theorem 2.4. If a rational function $F(z)$ has two decompositions

$$F_1 \circ F_2 \circ \ldots \circ F_r = G_1 \circ G_2 \circ \ldots \circ G_s$$

for which the conclusion of the first Ritt theorem does not hold then the algebraic curve corresponding to the equation

$$(F_1 \circ F_2 \circ \ldots \circ F_{r-1})(x) - (G_1 \circ G_2 \circ \ldots \circ G_{r-1})(y) = 0$$

is necessarily reducible.

In the second part of this paper, using the results of the first part, we construct explicitly a decomposition theory for the class $\mathcal{R}_2$. The reason for the investigation of this problem is twofold. On the one hand, this is a natural generalization of the Ritt theory. On the other hand, the decompositions of polynomials play an important role in the polynomial moment problem (see [13], [3]) which arose recently in connection with the "model" problem for the Poincare center-focus problem. 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 a decomposition theory for $\mathcal{R}_2$ can be considered as a preliminary step in the investigation of this problem.

Clearly, the description of double decompositions (3) of functions $h \in \mathcal{R}_2$ is essentially equivalent to the corresponding problem for Laurent polynomials and, since a Laurent polynomial has at most two poles, any such decomposition is equivalent to one of the following three decompositions:

$$A \circ L_1 = L_2 \circ z^d, \tag{4}$$

where $A(z)$ is a polynomial and $L_1(z), L_2(z)$ are Laurent polynomials,

$$A \circ L_1 = B \circ L_2 \tag{5}$$

where $A(z), B(z)$ are polynomials and $L_1(z), L_2(z)$ are Laurent polynomials, and

$$L_1 \circ z^{d_1} = L_2 \circ z^{d_2}, \tag{6}$$

where $L_1(z), L_2(z)$ are Laurent polynomials.

It is easy to see however that equality (6) implies that

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

for some Laurent polynomial $L(z)$ and $D = LCM(d_1, d_2)$. Furthermore, using Corollary 2.5 and some reasonings involving symmetries of the sphere, we show (Theorem 3.1) that any decompositions (4) is related either to a decomposition

$$z^n \circ z^r L(z^n) = z^r L^n(z) \circ z^n,$$

or to a decomposition

$$T_n \circ \frac{1}{2}(z^m + \frac{1}{z^m}) = \frac{1}{2}(z^m + \frac{1}{z^m}) \circ z^n,$$

where $L(z)$ is a Laurent polynomial and $T_n(z)$ is the nth Chebyshev polynomial.

Finally, the description of solutions of equation (5) is equivalent to the problem of description of curves (2), having a factor of genus 0 with at most two points at infinity, together with the corresponding parameterizations. Although the results of Fried and of Bilu and Tichy cited above reduce this problem solely to the finding of the corresponding parameterizations, in view of the great importance of equation (5) we provide an independent treatment of this equation since we believe that our method contains some new ideas which permit to simplify and clarify the existent approach to the problem.

Our analysis of equation (5) splits into three parts. In subsection 3.2 we describe solutions of (5) in the case when $\deg A(z) = \deg B(z)$. Further, in subsection 3.3 using this description we reduce the general case to the one when the pair $A(z)$, $B(z)$ is irreducible. Finally, in subsection 3.4 we solve (5) in the case when the pair $A(z), B(z)$ is irreducible. Here we propose a version of the formula for the genus g of curve (2) which permits to analyse the condition $g = 0$ in a convenient way and allows us to replace the conception of "extra" points which goes back to Ritt to a more transparent notion.

Eventually, in the end of the paper as a direct application of the classification of double decompositions and Theorem 2.5 we show that the first Ritt theorem extends to the class $\mathcal{R}_2$. The results of the second part of the paper can be summarized in the form of the following theorem which absorbs in particular the Ritt theorems and the Bilu-Tichy theorem.

Theorem 1.1 *Let*

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

be a double decomposition of a rational function $L \in \mathcal{R}_2$. Then either $A \circ C$ is equivalent to $B \circ D$ or there exist rational functions U, W, $\tilde{A}$, $\tilde{B} \in R_2$ such that

i) $$A = U \circ \tilde{A}, \quad B = U \circ \tilde{B}, \quad C = \tilde{C} \circ W, \quad D = \tilde{D} \circ W,$$

ii) $$\tilde{A} \circ \tilde{C} = \tilde{B} \circ \tilde{D},$$

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

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

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

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

where $S(z)$ 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 $T_n(z), T_m(z)$ *are Chebyshev polynomials,* $m,n\geq 1$, *and* $\mathrm{GCD}(n,m)=1$,

4) $$\tilde{A}\circ\tilde{C}\sim T_n\circ\frac{1}{2}(z^m+\frac{1}{z^m}), \qquad \tilde{B}\circ\tilde{D}\sim\frac{1}{2}(z^m+\frac{1}{z^m})\circ z^n,$$
where $m,n\geq 1$ *and* $\mathrm{GCD}(n,m)=1$,

5) $$\tilde{A}\circ\tilde{C}\sim -T_{nl}\circ\frac{1}{2}(\varepsilon z^m+\frac{\bar{\varepsilon}}{z^m}), \qquad \tilde{B}\circ\tilde{D}\sim T_{ml}\circ\frac{1}{2}(z^n+\frac{1}{z^n}),$$
where $T_{nl}(z), T_{ml}(z)$ *are Chebyshev polynomials,* $m,n\geq 1$, $l>1$, $\varepsilon^{nl}=-1$, *and* $\mathrm{GCD}(n,m)=1$,

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

Furthermore, if $\mathcal{D},\mathcal{E}$ *are two maximal decompositions of* L *then there exists a chain of decompositions* $\mathcal{F}_i$, $1\leq i\leq s$, *of* L *such that* $\mathcal{F}_1=\mathcal{D}$, $\mathcal{F}_s\sim\mathcal{E}$, *and* $\mathcal{F}_{i+1}$ *is obtained from* $\mathcal{F}_i$, $1\leq i\leq s-1$, *by a change of a segment of* $\mathcal{F}_i$ *consisting of two consecutive terms* $A\circ C$ *to a new segment* $B\circ D$ *such that* $A\circ C=B\circ D$.

Acknowledgments. This paper was written mostly during the visits of the author to the Max-Planck-Institut für Mathematik in Summer 2005 and Spring 2007 and 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 the preliminary results of this paper.

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

2.1. **Fundamental system of solutions.** In this subsection we establish some general properties of the functional equation $h=f\circ p=g\circ q$, where $f: C_1\rightarrow\mathbb{CP}^1$, $g: C_2\rightarrow\mathbb{CP}^1$ are fixed holomorphic functions on fixed Riemann surfaces C_1, C_2 and $h: C\rightarrow\mathbb{CP}^1$, $p: C\rightarrow C_1$, $q: C\rightarrow C_2$ are unknown holomorphic functions on unknown Riemann surface C.

Let $S\subset\mathbb{CP}^1$ be a union of branch points of f and g and let z_0 be a point from $\mathbb{CP}^1\setminus S$. Recall that for any collection consisting of a Riemann surface R, holomorphic function $p: R\rightarrow\mathbb{CP}^1$ non ramified outside of S, and a point $e\in p^{-1}\{z_0\}$ the homomorphism of the fundamental groups
$$p_\star:\ \pi_1(R\setminus p^{-1}\{S\},e)\rightarrow\pi_1(\mathbb{CP}^1\setminus S,z_0)$$
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 $p: R\rightarrow\mathbb{CP}^1$, and a point $e\in p^{-1}\{z_0\}$ such that
$$p_\star(\pi_1(R\setminus p^{-1}\{S\},e))=\Gamma.$$
Furthermore, this correspondence descends to a one-to-one correspondence between conjugacy classes of subgroup 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 $p_1 : R_1 \to \mathbb{CP}^1$, $e_1 \in p_1^{-1}\{z_0\}$ 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 general, 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 case if such a function exists it is defined in a unique way.

In view of the fact that the coverings of Riemann surfaces are identified with holomorphic functions the results above follow from the corresponding results about coverings (see e.g. [11]).

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 a holomorphic function $h : C \to \mathbb{CP}^1$, a point $c \in h^{-1}\{z_0\}$, and holomorphic functions $u : C \to C_1$, $v : C \to C_2$ such that*

$$h = f \circ u = g \circ v, \quad u(c) = a, \quad v(c) = b. \tag{7}$$

Furthermore, the function $h(z)$ has the following property: if

$$\tilde{h} = f \circ p = g \circ q, \quad p(\tilde{c}) = a, \quad q(\tilde{c}) = b \tag{8}$$

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

$$\tilde{h} = h \circ \tilde{w}, \quad p = u \circ \tilde{w}, \quad q = v \circ \tilde{w}, \quad \tilde{w}(\tilde{c}) = c. \tag{9}$$

Proof. Indeed, it is easy to see that for the pair $h : C \to \mathbb{CP}^1$, $c \in h^{-1}\{z_0\}$ corresponding to the subgroup $\Gamma_{f,a} \cap \Gamma_{g,b}$ equalities (7) hold. Furthermore, if equalities (8) hold then $\Gamma_{\tilde{h},\tilde{c}} \subseteq \Gamma_{h,c}$ and therefore $\tilde{h} = h \circ \tilde{w}$ for some $\tilde{w} : R \to C$, such that $\tilde{w}(\tilde{c}) = c$. Since

$$(u \circ \tilde{w})(\tilde{c}) = a = p(\tilde{c}), \qquad (v \circ \tilde{w})(\tilde{c}) = b = q(\tilde{c})$$

we conclude that

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

Fix a numeration $\{z_1, z_2, \dots, z_r\}$ of points of S and let $h : R \to \mathbb{CP}^1$ be a holomorphic function non ramified outside of S. For each i, $1 \le i \le r$, a small loop around β_i after lifting by $h(z)$ induces a permutation $\alpha_i(h)$ of points of $h^{-1}\{z_0\}$. Furthermore, the equality $\alpha_1(h)\alpha_2(h)\dots\alpha_r(h) = 1$ holds and the group G_h generated by $\alpha_i(h)$, $1 \le i \le r$, is transitive. The group G_h is called the monodromy group of $p(z)$. Clearly, the representation of $\alpha_i(h)$, $1 \le i \le r$, by elements of the symmetric group S_d depends on the numeration of points of $h^{-1}\{z_0\}$ 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, $1 \le i \le r$, from S_d generating a transitive permutation group and such that $\alpha_1\alpha_2\dots\alpha_r = 1$ (see e.g. [12], Corollary 4.10). We will denote the conjugacy class corresponding to the function $h(z)$ by $\hat{\alpha}(h)$.

Notice that if

$$\varphi_h : \pi_1(\mathbb{CP}^1 \setminus S, z_0) \to G_h$$

is a homomorphism which sends β_i to α_i, $1 \le i \le r$, then the preimages of the stabilizers of the elements of $\{1, 2, \dots, d\}$ in G_h coincide with the groups $\Gamma_{h,e}$, $e \in h^{-1}\{z_0\}$. On the other hand, if a group $\Gamma_{h,e}$ for some $e \in h^{-1}\{z_0\}$ is given then the collection of permutations α_i, $1 \le i \le r$, induced on the cosets of $\Gamma_{h,e}$ by β_i, $1 \le i \le r$, is a representative of $\hat{\alpha}(h)$.

Recall that if $h = f \circ p$ is a decomposition of a holomorphic function $h : R \to \mathbb{CP}^1$ of degree d into a composition of holomorphic functions $p : R \to C$ and $f : C \to \mathbb{CP}^1$ then the group G_h has an imprimitivity system Ω such that the collection of permutations of blocks of Ω induced by $\alpha_i(h)$, $1 \le i \le r$, is a representative of $\hat{\alpha}(f)$. Namely, after an identification of the set $h^{-1}\{z_0\}$ with the set $\{1, 2, \dots, d\}$ the blocks of Ω coincide with the preimages $p^{-1}\{t\}$ of the points $t \in f^{-1}\{z_0\}$. Furthermore, two different decompositions $h = f \circ p$ and $h = \tilde{f} \circ \tilde{p}$, where $\tilde{f} : \tilde{C} \to \mathbb{CP}^1$ and $\tilde{p} : R \to \tilde{C}$ lead to the same imprimitivity system if and only there exists an automorphism $\mu : \tilde{C} \to C$ such that

$$f = \tilde{f} \circ \mu^{-1}, \quad p = \mu \circ \tilde{p}.$$

We will call such decompositions equivalent. Vice versa, if G_h has an imprimitivity system Ω such that the collection of permutations of blocks of Ω induced by $\alpha_i(h)$, $1 \le i \le r$, is a representative of $\hat{\alpha}(f)$ then there exists a function $p : R \to C$ such that $h = f \circ p$. Therefore, non-equivalent decompositions of $h(z)$ are in a one-to-one correspondence with imprimitivity systems of G_h.

Notice that the blocks of two imprimitivity systems Ω_1 and Ω_2, corresponding to decompositions $h = f_1 \circ p_1$ and $h = f_2 \circ p_2$ for some $f_1 : C_1 \to \mathbb{CP}^1$, $f_2 : C_2 \to \mathbb{CP}^1$, $p_1 : R \to C_1$, $p_2 : R \to C_2$, and containing a common element, have an intersection of the cardinality d if and only if there exist a holomorphic function $w : R \to \tilde{R}$ of degree d and holomorphic functions $p_1 : \tilde{R} \to C_1$, $p_2 : \tilde{R} \to C_2$ such that

$$p_1 = \tilde{p}_1 \circ w, \qquad p_2 = \tilde{p}_2 \circ w.$$

In particular, the function $p(z)$ in a decomposition $h = f \circ p$ is defined by the corresponding imprimitivity system and a choice of $f(z)$ in a unique way up to a composition $\omega \circ p$ with an automorphism ω of the surface C such that $f \circ \omega = f$.

Notice also that the set of all blocks of G_h containing a fixed element i is in a one-to-one correspondence with the subgroups G of G_h containing the stabilizer $G_{h,i}$ of i. Namely, if G is such a group then its orbit containing i is a block. In particular, a function $h(z)$ is indecomposable if and only if $G_{h,i}$ is a maximal subgroup of G_h.

Let $f : C_1 \to \mathbb{CP}^1$ be a holomorphic function of degree n and $g : C_2 \to \mathbb{CP}^1$ be a holomorphic function of degree m. Fix some representatives $\alpha_i(f)$, $\alpha_i(g)$, $1 \le i \le s$, of the classes $\hat{\alpha}(f)$, $\hat{\alpha}(g)$ and define permutations $\delta_1, \delta_2, \dots, \delta_r \in S_{nm}$ as follows: consider the set of mn elements c_{j_1,j_2}, $1 \le j_1 \le n$, $1 \le j_2 \le m$, and set $c_{j_1,j_2}^{\delta_i} = c_{j_1',j_2'}$, where

$$j_1' = j_1^{\alpha_i(f)}, \quad j_2' = j_2^{\alpha_i(g)}, \quad 1 \le i \le s.$$

It is convenient to consider c_{j_1,j_2}, $1 \le j_1 \le n$, $1 \le j_2 \le m$, as elements of a $n \times m$ matrix M. Then the action of the permutation δ_i, $1 \le i \le r$, 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, $1 \leq i \leq r$, is not transitive on the set c_{j_1,j_2}, $1 \leq j_1 \leq n$, $1 \leq j_2 \leq m$. Denote by $o(f,g)$ the number of transitivity sets of the group $\Gamma(f,g)$ and let $\delta_i(j)$, $1 \leq j \leq o(f,g)$, $1 \leq i \leq r$, be a permutation induced by the permutations δ_i, $1 \leq i \leq r$, on the transitivity set U_j, $1 \leq j \leq o(f,g)$. Since for any j, $1 \leq j \leq o(f,g)$, the equality

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

holds there exist holomorphic functions $h_j : R_j \rightarrow \mathbb{CP}^1$, $1 \leq j \leq o(f,g)$, such that the collection $\delta_i(j)$, $1 \leq i \leq r$, is a representative of $\hat{\alpha}(h_j)$.

By construction the group G_j generating by $\delta_i(j)$, $1 \leq i \leq s$, is transitive and has two imprimitivity systems $\Omega_f(j), \Omega_g(j)$ such that the permutation of blocks of $\Omega_f(j)$ (resp. of $\Omega_g(j)$) induced by $\delta_i(j)$, $1 \leq i \leq r$, is a representative of $\hat{\alpha}(f)$ (resp. of $\hat{\alpha}(g)$). Therefore, there exist holomorphic functions $u_j : R_j \rightarrow C_1$ and $v_j : R_j \rightarrow C_2$ such that

$$h_j = f \circ u_j = g \circ v_j. \tag{10}$$

Theorem 2.1. *Let $f : C_1 \rightarrow \mathbb{CP}^1$, $g : C_2 \rightarrow \mathbb{CP}^1$ be holomorphic functions. Suppose that $h : R \rightarrow \mathbb{CP}^1$ is a holomorphic function such that*

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

for some holomorphic functions $p : R \rightarrow C_1$ and $q : R \rightarrow C_2$. Then there exists a number j, $1 \leq j \leq o(f,g)$, and holomorphic functions $w : R \rightarrow R_j$, $\tilde{p} : R_j \rightarrow C_1$, $\tilde{q} : R_j \rightarrow C_2$, such that

$$h = h_j \circ w, \quad p = \tilde{p} \circ w, \quad q = \tilde{q} \circ w \tag{12}$$

and

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

Proof. It is enough to prove that for any choice of points $a \in f^{-1}\{z_0\}$ and $b \in g^{-1}\{z_0\}$ the class of permutations $\hat{\alpha}(h)$ for the corresponding function $h(z)$ from Proposition 2.1 coincides with $\hat{\alpha}(h_j)$ for some j, $1 \leq j \leq o(f,g)$. On the other hand, the last statement is equivalent to the statement that for any $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ there exists j, $1 \leq j \leq o(f,g)$, 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

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

Let l be a number 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 a number 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 a transitivity set for the group $\Gamma(f,g)$ containing the element $c_{l,k}$. We have:

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

Furthermore, if

$$\psi_1 : G_f \rightarrow G_j, \quad \psi_2 : G_g \rightarrow G_j$$

are homomorphisms which send $\alpha_i(f)$, $1 \leq i \leq r$, (resp. $\alpha_i(g)$, $1 \leq i \leq r$,) to $\alpha_i(h_j)$, $1 \leq i \leq r$, then

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

where A_l (resp. B_k) is a 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.

Observe now that

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

Therefore,

$$\Gamma_{f,a} \cap \Gamma_{g,b} = (\psi_1 \circ \varphi_f)^{-1}\{A_l\} \cap (\psi_2 \circ \varphi_g)^{-1}\{B_k\} =$$

$$= \varphi_{h_j}^{-1}\{A_l\} \cap \varphi_{h_j}^{-1}\{B_k\} = \varphi_{h_j}^{-1}\{A_l \cap B_k\} = \varphi_{h_j}^{-1}\{G_{j,c}\}. \qquad \square$$

Let $g(R_j)$ be the genus of the surface R_j, $1 \leq j \leq o(f,g)$, on which the function $h_j(z)$ is defined. Notice that Proposition 2.1 implies in particular that if $h : R \to \mathbb{CP}^1$ is a holomorphic function which satisfies equation (11) for some functions $p : R \to C_1$ and $q : R \to C_2$ then necessary $g(R) \geq \min_j g(R_j)$. In particular, if $\min_j g(R_j) > 0$ then (11) does not have rational solutions.

Denote by

$$\lambda_1 = (f_{1,1}, f_{1,2}, ..., f_{1,u_1}), \quad ... \quad , \lambda_r = (f_{r,1}, f_{r,2}, ..., f_{r,u_r})$$

and

$$\mu_1 = (g_{1,1}, g_{1,2}, ..., g_{1,v_1}), \quad ... \quad , \mu_s = (g_{r,1}, g_{r,2}, ..., g_{r,v_r})$$

the collections of partitions of the numbers $n = \deg f(z)$ and $m = \deg g(z)$ corresponding to the decompositions of the permutations $\alpha_i(f)$, $1 \leq i \leq r$, and $\alpha_i(g)$, $1 \leq i \leq r$, into products of disjoint cycles.

Proposition 2.2. *The formula*

$$\sum_{i=1}^{r} \sum_{j_1=1}^{u_i} \sum_{j_2=1}^{v_i} \mathrm{GCD}(f_{i,j_1} g_{i,j_2}) = (r-2)nm + 2 - 2 \sum_{j=1}^{o(f,g)} g(R_j) \tag{13}$$

holds.

Proof. Indeed, if

$$\nu_1(j) = (c_{1,1}(j), c_{1,2}(j), ..., c_{1,e_1(j)}(j)), \ ... \ , \nu_r(j) = (c_{r,1}(j), c_{r,2}(j), ..., c_{r,e_r(j)}(j))$$

is a collection of partitions of the number $|U_j|$, $1 \leq j \leq o(f,g)$, corresponding to decompositions of the permutations $\alpha_i(h_j)$, $1 \leq i \leq r$, into products of disjoint cycle then by the Riemann-Hurwitz formula we have:

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

and therefore

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

On the other hand, it follows from the construction that the permutation δ_i, $1 \leq i \leq r$, contains

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

disjointed cycles and therefore

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

2.2. **Irreducible and reducible pairs.** It follows from Theorem 2.1 that solutions of (11) have especially simple form in the case when the group $\Gamma(f,g)$ is transitive on the set c_{j_1,j_2}, $1 \le j_1 \le n$, $1 \le j_2 \le m$. In this case say that the pair of functions f and g is *irreducible* otherwise say that it is *reducible*. In this subsection we study properties of irreducible and reducible pairs.

Proposition 2.3. *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 $\Gamma_{f,a} \cap \Gamma_{g,b}$ coincides with the cardinality of the corresponding imprimitivity set U_j it is clear that 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

$$[\pi_1(\mathbb{CP}^1 \setminus S, z_0) : \Gamma_{f,a} \cap \Gamma_{g,b}] = nm \tag{14}$$

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 equals mn whenever n and m are coprime.

Theorem 2.2. *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*

$$\Gamma_{f,a}\Gamma_{g,b} = \Gamma_{g,b}\Gamma_{f,a} = \pi_1(\mathbb{CP}^1 \setminus S, z_0) \tag{15}$$

holds.

Proof. Since

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

the equality (14) is equivalent to the equality

$$[\Gamma_{g,b} : \Gamma_{f,a} \cap \Gamma_{g,b}] = n. \tag{16}$$

On the other hand, for any subgroups A, B of finite index in a group G the inequality

$$[\{A, B\} : A] \ge [B : A \cap B]$$

holds and the equality attains if and only if the groups A and B commute (see e.g. [9], p. 79). Therefore,

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

and hence equality (16) holds if and only if $\Gamma_{f,a}$ and $\Gamma_{g,b}$ are permutable and equality (15) holds. $\square$

Corollary 2.1. *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. Indeed, it follows from the inclusions

$$\Gamma_{f,a} \subset \Gamma_{\tilde{f},\tilde{a}}, \quad \Gamma_{g,b} \subset \Gamma_{\tilde{g},\tilde{b}}$$

for the corresponding subgroups that

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

Recall that a holomorphic function $f : C_1 \to \mathbb{CP}^1$ is called normal if the subgroup of $\mathrm{Aut}\,(C_1)$ consisting from the automorphisms σ for which $f \circ \sigma = f$ acts transitively on the set $f^{-1}\{z_0\}$. An equivalent condition is that for any $a, b \in f^{-1}\{z_0\}$ the equality $\Gamma_{f,a} = \Gamma_{g,b}$ holds. For a function $f : C_1 \to \mathbb{CP}^1$ denote by $N(f)$ its normalization that is a normal function which corresponds to the group

$$\bigcap_{a \in f^{-1}\{z_0\}} \Gamma_{f,a}.$$

Theorem 2.3. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_2 \to \mathbb{CP}^1$ be holomorphic functions. Suppose that the pair f, g is reducible. Then there exist holomorphic functions $f_1 : \tilde{C}_1 \to \mathbb{CP}^1$, $g_1 : \tilde{C}_2 \to \mathbb{CP}^1$, and $u : C_1 \to \tilde{C}_1$, $v : C_2 \to \tilde{C}_2$ such that*

$$f = f_1 \circ u, \quad g = g_1 \circ v, \quad \textit{and} \quad N(f_1) = N(g_1). \tag{17}$$

Proof. Without loss of generality we will assume that for any $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ the group $C = \{\Gamma_{f,a}, \Gamma_{g,b}\}$ coincides with the group $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ since otherwise $C = \Gamma_{h,c}$ for some $h : C \to \mathbb{CP}^1$ and $c \in h^{-1}\{z_0\}$ such that

$$f = h \circ u, \quad g = h \circ v$$

for some $u : C_1 \to C$, $v : C_2 \to C$, and we can set $f_1 = g_1 = h$.

Let $a \in f^{-1}\{z_0\}$, $b \in g^{-1}\{z_0\}$ and let $d(f)$ (resp. $d(g)$) be a maximal number such that there exists a chains of subgroups

$$\Gamma_{f,a} = K_0 \subset K_2 \subset \dots \subset K_{d(f)} = \pi_1(\mathbb{CP}^1 \setminus S, z_0)$$

(resp.

$$\Gamma_{g,b} = L_0 \subset L_2 \subset \dots \subset L_{d(g)} = \pi_1(\mathbb{CP}^1 \setminus S, z_0)\ ).$$

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

Consider first the case when $d = 2$ that is when both functions f, g are indecomposable. Since $N(g)$ is a normal subgroup of $\pi_1(\mathbb{CP}^1 \setminus S, z_0)$ we have:

$$\{\Gamma_{f,a}, N(g)\} = \Gamma_{f,a}N(g) = N(g)\Gamma_{f,a}. \tag{18}$$

Furthermore, the equality $d(f) = 1$ implies that either

$$\{\Gamma_{f,a}, N(g)\} = \Gamma_{f,a} \tag{19}$$

or

$$\{\Gamma_{f,a}, N(g)\} = \pi_1(\mathbb{CP}^1 \setminus S, z_0). \tag{20}$$

The last equality however would imply that

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

in contradiction with Proposition 2.2. Therefore, equality (19) holds and hence $N(g) \subseteq \Gamma_{f,a}$. Since this inclusion holds for any $a \in f^{-1}\{z_0\}$ we conclude that

$$N(g) \subseteq \bigcap_{a \in f^{-1}\{z_0\}} \Gamma_{f,a} = N(f).$$

The same arguments show that $N(f) \subseteq N(g)$. Therefore, $N(g) = N(f)$.

Suppose now that the proposition is proved for all pairs with $d < n$ and let f, g be a pair with $d = n$. If $N(f) = N(g)$ then we can set $f_1 = f$, $g_1 = g$ so assume that $N(f) \neq N(g)$. Then either there exists $a \in f^{-1}\{z_0\}$ such that

$$\Gamma_{f,a} \subset \{\Gamma_{f,a}, N(g)\},$$

or there there exists $b \in g^{-1}\{z_0\}$ such that

$$\Gamma_{g,b} \subset \{\Gamma_{g,b}, N(f)\}.$$

Suppose that say $\Gamma_{f,a}$ is a proper subgroup of the group $G = \{\Gamma_{f,a}, N(g)\}$ for some $a \in f^{-1}\{z_0\}$. Since equality (20) is impossible this implies that there exist $h : C \to \mathbb{CP}^1$ and $c \in h^{-1}\{z_0\}$ such that $G = \Gamma_{h,c}$.

Observe that the groups $\Gamma_{h,c}$ and $\Gamma_{g,b}$ do not commute since otherwise in view of equality (18) we would have:

$$\Gamma_{f,a}\Gamma_{g,b} = \Gamma_{f,a}(N(g)\Gamma_{g,b}) = (\Gamma_{f,a}N(g))\Gamma_{g,b} = \Gamma_{g,b}(\Gamma_{f,a}N(g)) =$$
$$= \Gamma_{g,b}(N(g)\Gamma_{f,a}) = (\Gamma_{g,b}(N(g))\Gamma_{f,a} = \Gamma_{g,b}\Gamma_{f,a}.$$

Therefore, the pair h, g is reducible. Since by construction $f = h \circ p$ for some $p : C_1 \to C$ and $\deg h(z) < \deg f(z)$, the proposition follows now from the induction assumption. □

Theorem 2.4. *Let $f : C_1 \to \mathbb{CP}^1$, $g : C_g \to \mathbb{CP}^1$ be an irreducible pair of holomorphic functions and $p : C \to C_1$, $q : C \to C_2$ be indecomposable holomorphic functions such that $f \circ p = g \circ q$. Then $f(z)$ and $g(z)$ are indecomposable.*

Proof. Fix a point $c \in h^{-1}\{z_0\}$, where $h = f \circ p = g \circ q$. To the decompositions $f \circ p$ and $g \circ q$ correspond the inclusions

$$\Gamma_{h,c} \subset \Gamma_{f,x_1} \qquad \Gamma_{h,c} \subset \Gamma_{g,x_2},$$

where $x_1 = p(c), x_2 = q(c)$. Furthermore, it follows from the indecomposability of $p(z)$ and $q(z)$ that

$$\Gamma_{h,c} = \Gamma_{f,x_1} \cap \Gamma_{g,x_2}. \tag{21}$$

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

$$\Gamma_{f,x_1} \subset \Gamma \tag{22}$$

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

Consider the intersection

$$\Gamma_1 = \Gamma \cap \Gamma_{g,x_2}.$$

By Theorem 2.2

$$\Gamma_{f,x_1}\Gamma_{g,x_2} = \pi_1(\mathbb{CP}^1 \setminus S, z_0).$$

Therefore,

$$\Gamma\,\Gamma_{g,x_2} = \pi_1(\mathbb{CP}^1 \setminus S, z_0).$$

By the statement cited in Theorem 2.2 we have:

$$[\pi_1(\mathbb{CP}^1 \setminus S, z_0) : \Gamma_{f,x_1}] = [\Gamma_{g,x_2} : \Gamma_{h,c}], \quad [\pi_1(\mathbb{CP}^1 \setminus S, z_0) : \Gamma] = [\Gamma_{g,x_2} : \Gamma_1].$$

Therefore, (22) implies that $\Gamma_{h,c} \subset \Gamma_1$. Since $\Gamma_1 \subseteq \Gamma_{g,x_2}$ it follows now from the indecomposability of $q(z)$ that $\Gamma_1 = \Gamma_{g,x_2}$. Therefore, $\Gamma_{f,x_1}\Gamma_{g,x_2} \subseteq \Gamma$ and hence $\Gamma = \pi_1(\mathbb{CP}^1 \setminus S, z_0)$. □

Notice implicitly the following corollary of Theorem 2.4.

Corollary 2.2. *Suppose that a rational function $f(z)$ has two maximal decompositions*

$$f_r \circ f_{r-1} \circ \ldots \circ f_1 = g_s \circ g_{s-1} \circ \ldots \circ g_1$$

for which the conclusion of the first Ritt theorem does not hold. Set

$$f_r \circ f_{r-1} \circ \ldots \circ f_2 = \frac{A_1(z)}{B_1(z)}, \qquad g_s \circ g_{s-1} \circ \ldots \circ g_2 = \frac{A_2(z)}{B_2(z)},$$

where $A_1(z)$, $B_1(z)$ and $A_2(z)$, $B_2(z)$ are pairs of polynomials with no common roots. Then the algebraic curve

$$A_1(x)B_2(y) - A_2(y)B_1(x) = 0$$

is reducible. $\square$

2.3. **Double decompositions involving generalized polynomials.** Say that a holomorphic function $h : C \rightarrow \mathbb{CP}^1$ is a generalized polynomial if $h^{-1}\{\infty\}$ consists of a unique point. The double decompositions $f \circ p = g \circ q$ for which $f(z), g(z)$ are generalized polynomials have a number of special properties.

Corollary 2.3. *If, in notation of Theorem 2.3, the functions $f(z), g(z)$ are generalized polynomials then* $\deg f_1 = \deg g_1$.

Proof. Indeed, the group corresponding to $N(f_1)$ can obtained by a consecutive intersection of the groups $\Gamma_{f_1,a}$, $a \in f_1^{-1}\{z_0\}$. Since f_1 is a generalized polynomial, it is easy to see that at each stage the permutation corresponding to the loop around infinity of the corresponding permutation group consists of cycles of length equals to the degree f_1 only. Therefore, the same is true for $N(f_1)$. Hence $N(f_1) = N(g_1)$ implies that $\deg f_1 = \deg g_1$. $\square$

Proposition 2.4. *Let $f : C_1 \rightarrow \mathbb{CP}^1$, $g : C_2 \rightarrow \mathbb{CP}^1$ be generalized polynomials, $n = \deg f(z)$, $m = \deg g(z)$. Set $y_f = f^{-1}\{\infty\}, y_g = g^{-1}\{\infty\}$, $l = \mathrm{LCM}(n, m)$. Suppose that there exist holomorphic functions $p : R \rightarrow C_1$, $q : R \rightarrow C_2$ such that*

$$f \circ p = g \circ q. \tag{23}$$

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

$$p = \tilde{p} \circ w, \quad q = \tilde{q} \circ w \tag{24}$$

and the following property holds: the multiplicity of any point from $\tilde{p}^{-1}\{y_f\}$ equals l/n while the multiplicity of any point from $\tilde{q}^{-1}\{y_g\}$ equals l/m.

Proof. Set $h = f \circ p = g \circ q$. Let $\mathfrak{I}$ (resp. $\mathfrak{J}$) be an imprimitivity system of the group $G_h \subseteq S_n$ corresponding to the decomposition $h = f \circ p$ (resp. $h = g \circ q$) and $\mathfrak{I}_1$ (resp. $\mathfrak{J}_1$) be a block of $\mathfrak{I}$ (resp. $\mathfrak{J}$) containing 1. The equality (24) holds for some function $w(z)$ of degree greater than one if and only if $|\mathfrak{I}_1 \cap \mathfrak{J}_1| > 1$, and without loss of generality we can assume that $|\mathfrak{I}_1 \cap \mathfrak{J}_1| = 1$. Therefore, in order to prove the theorem it is enough to show that if the multiplicity k of some point from $h^{-1}\{\infty\}$ with respect to $h(z)$ is greater than l then $|\mathfrak{I}_1 \cap \mathfrak{J}_1| > 1$.

Let $\sigma \in G_h$ be a permutation corresponding to the loop around infinity. Consider any cycle o from the decomposition of σ into a product of disjointed cycles. Without loss of generality we can assume that $o = (1, 2, ..., k)$. Since $f(z)$ is a generalized polynomial it is easy to see that the intersection of o with $\mathfrak{I}_1$ consists of numbers congruent to 1 by modulo n. Similarly, the intersection of o with $\mathfrak{J}_1$ consists of numbers congruent to 1 by modulo m. Therefore, if $k > l$ then $|\mathfrak{I}_1 \cap \mathfrak{J}_1| > 1$. $\square$

Corollary 2.4. *Suppose that under assumptions of Proposition 2.4 the function $h = f \circ p = g \circ q$ is a generalized polynomial and $\deg f(z) = \deg g(z)$. Then $f \circ p \sim g \circ q$.*

Proof. Indeed, in this case the set $\tilde{p}^{-1}\{y_f\}$ contains a unique point and the multiplicity of this point with respect to $\tilde{p}(z)$ is one. Therefore $\tilde{p}(z)$ is an automorphism. The same is true for $\tilde{q}(z)$. $\square$

Let us mention explicitly the following specification of Proposition 2.4 which we will use in the following.

Corollary 2.5. *Let $A(z)$, $B(z)$ be polynomials of the same degree and $C(z)$, $D(z)$ be rational functions such that*

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

Then there exist rational functions $\tilde{C}(z)$, $\tilde{D}(z)$ such that

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

and $\tilde{C}(z)$, $\tilde{D}(z)$ have equal number of poles all of which are simple. $\square$

2.4. **Ritt classes of rational functions.** Say that a family of rational functions H is *a closed class* if for any $f \in H$ the equality $f = g \circ h$ implies that $g \in H$, $h \in H$. For example, rational functions for which

$$\min_{z \in \mathbb{CP}^1} \sharp\{f^{-1}\{z\}\} \leq k,$$

where $k \geq 1$ is a fixed number, form a closed class which we will denote by $\mathcal{R}_k$.

Say that two maximal decompositions $\mathcal{D}, \mathcal{E}$ of a rational function $f(z)$ are weakly equivalent if there exists a chain of decompositions $\mathcal{F}_i$, $1 \leq i \leq s$, of $f(z)$ such that $\mathcal{F}_1 = \mathcal{D}$, $\mathcal{F}_s \sim \mathcal{E}$, and $\mathcal{F}_{i+1}$ is obtained from $\mathcal{F}_i$, $1 \leq i \leq s-1$, by a change of a segment of $\mathcal{F}_i$ consisting of two consecutive terms $A \circ B$ to a new segment $C \circ D$ such that $A \circ C = B \circ D$. It is easy to see that this is indeed an equivalence relation which we will denote by the symbol $\sim_w$. Say that a closed class of rational functions H is *the Ritt class* if any two maximal decompositions of any function $f(z)$ from H are weakly equivalent.

Finally, say that a double decomposition $f \circ p = g \circ q$ is elementary if $p(z)$, $q(z)$ are indecomposable and there exist no rational functions $\tilde{f}(z)$, $\tilde{g}(z)$, $u(z)$ with $\deg u(z) > 1$ such that

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

and

$$\tilde{f} \circ p = \tilde{g} \circ q.$$

Theorem 2.5. *Let H be a closed class of rational functions. Suppose that for any function $h \in H$ and any elementary double decomposition*

$$h = f \circ p = g \circ q$$

such that the pair $f(z), g(z)$ is reducible, for any choice of maximal decompositions

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

the decompositions

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

are weakly equivalent. Then H is the Ritt class.

Proof. We use the induction on the degree of $h(z)$. If $\deg h(z) = 1$ then the conclusion of the theorem is true.

Suppose now that $\deg h(z) > 1$ and let

$$\mathcal{H}_1 : \ h = f_r \circ f_{r-1} \circ \dots \circ f_1, \qquad \mathcal{H}_2 : \ h = g_s \circ g_{s-1} \circ \dots \circ g_1$$

be two decompositions of a function $h(z) \in H$. Set

$$f = f_r \circ f_{r-1} \circ \dots \circ f_2, \qquad g = g_s \circ g_{s-1} \circ \dots \circ g_2$$

and consider the double decomposition

$$f \circ f_1 = g \circ g_1.$$

If this decomposition is elementary then Theorem 2.4 and the condition of the theorem imply that $\mathcal{H}_1$ and $\mathcal{H}_2$ are weakly equivalent. Otherwise there exist rational functions $\tilde{f}(z)$, $\tilde{g}(z)$, $u(z)$, $\deg u(z) > 1$, such that

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

and the double decomposition

$$\tilde{f} \circ f_1 = \tilde{g} \circ g_1 \tag{25}$$

is elementary.

Let

$$u = u_l \circ u_{l-1} \circ \dots \circ u_1, \quad \tilde{f} = \tilde{f}_n \circ \tilde{f}_{n-1} \circ \dots \circ \tilde{f}_1, \quad \tilde{g} = \tilde{g}_m \circ \tilde{g}_{\tilde{m}-1} \circ \dots \circ \tilde{g}_1$$

be some maximal decompositions of the functions $u(z)$, $\tilde{f}(z)$, $\tilde{g}(z)$. By the induction assumption

$$f_r \circ f_{r-1} \circ \dots \circ f_2 \ \sim_w \ u_l \circ u_{l-1} \circ \dots \circ u_1 \circ \tilde{f}_n \circ \tilde{f}_{n-1} \circ \dots \circ \tilde{f}_1.$$

Therefore, also

$$f_r \circ f_{r-1} \circ \dots \circ f_2 \circ f_1 \ \sim_w \ u_l \circ u_{l-1} \circ \dots \circ u_1 \circ \tilde{f}_n \circ \tilde{f}_{n-1} \circ \dots \circ \tilde{f}_1 \circ f_1. \tag{26}$$

In a similar way we conclude that

$$g_s \circ g_{s-1} \circ \dots \circ g_2 \circ g_1 \ \sim_w \ u_l \circ u_{l-1} \circ \dots \circ u_1 \circ \tilde{g}_m \circ \tilde{g}_{\tilde{m}-1} \circ \dots \circ \tilde{g}_1 \circ g_1. \tag{27}$$

Since the double decomposition (25) is elementary it follows from (26), (27) taking into account Theorem 2.4 and the condition of the theorem that $\mathcal{H}_1 \sim_w \mathcal{H}_2$. $\Box$

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

Proposition 2.5. *The class $\mathcal{R}_1$ is the Ritt class.*

Proof. Clearly, in view of Theorem 2.5 it is enough to prove that if

$$f \circ p = g \circ q \tag{28}$$

is an elementary decomposition of a polynomial $h(z)$ then the pair $f(z)$, $g(z)$ is irreducible.

Assume the inverse. Then by Corollary 2.3 there exist polynomials $\tilde{f}(z)$, $\tilde{g}(z)$, $\deg \tilde{f} = \deg \tilde{g} > 1$, and rational functions $u(z)$, $v(z)$ such that

$$f = \tilde{f} \circ u, \quad g = \tilde{g} \circ v.$$

Since Corollary 2.4 implies that

$$\tilde{f} \circ (u \circ p) \sim \tilde{g} \circ (v \circ q)$$

we obtain a contradiction with the assumption that (28) is elementary. $\Box$

3. Decompositions of Laurent polynomials

3.1. Solutions of equation (4).

Lemma 3.1. *Let $U(z), V(z)$ be rational functions such that*

$$U \circ z^n = V \circ \frac{1}{2}(z + \frac{1}{z}), \quad n \geq 1. \tag{29}$$

Then there exists a rational function $R(z)$ such that

$$U = R \circ \frac{1}{2}(z + \frac{1}{z}), \qquad V = R \circ T_n.$$

Similarly, if

$$U \circ z^{d_1} = V \circ z^{d_2}, \quad d_1, d_2 \geq 1,$$

then there exists a rational function $R(z)$ such that

$$U = R \circ z^{D/d_1}, \qquad V = R \circ z^{D/d_2},$$

where $D = LCM(d_1, d_2)$.

Proof. Since the function $F(z)$ to which correspond decompositions (29) is invariant with respect to the automorphisms $\alpha_1 : z \to \varepsilon z$ and $\alpha_2 : z \to \frac{1}{z}$ it is invariant with respect to the group generated by α_1, α_2 and therefore

$$F = R \circ \frac{1}{2}(z^m + \frac{1}{z^m})$$

for some rational function $R(z)$. It follows now from

$$R \circ \frac{1}{2}(z^m + \frac{1}{z^m}) = R \circ \frac{1}{2}(z + \frac{1}{z}) \circ z^n = U \circ z^n,$$

and

$$R \circ \frac{1}{2}(z^m + \frac{1}{z^m}) = R \circ T_n \circ \frac{1}{2}(z + \frac{1}{z}) = V \circ \frac{1}{2}(z + \frac{1}{z}).$$

that

$$U = R \circ \frac{1}{2}(z + \frac{1}{z}), \qquad V = R \circ T_n.$$

The proof of the second part of the lemma is similar. □

Theorem 3.1. *Suppose that polynomials $A(z), D(z)$ and Laurent polynomials $L_1(z)$, $L_2(z)$ (which are not polynomials) satisfy the equation*

$$A \circ L_1 = L_2 \circ D. \tag{30}$$

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

i) $$A = P \circ \tilde{A}, \quad L_2 = P \circ \tilde{L}_2, \quad L_1 = \tilde{L}_1 \circ W, \quad D = \tilde{D} \circ W,$$

ii) $$\tilde{A} \circ \tilde{L}_1 = \tilde{L}_2 \circ \tilde{D},$$

and either

$$\tilde{A} \circ \tilde{L}_1 \sim z^n \circ z^r L(z^n), \qquad \tilde{L}_2 \circ \tilde{D} \sim z^r L^n(z) \circ z^n, \tag{31}$$

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

$$\tilde{A} \circ \tilde{L}_1 \sim T_n \circ \frac{1}{2}(z^m + \frac{1}{z^m}), \qquad \tilde{L}_2 \circ \tilde{D} \sim \frac{1}{2}(z^m + \frac{1}{z^m}) \circ z^n, \tag{32}$$

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

Proof. Without loss of generality we can assume that $D(z) = z^d$ for some $d > 1$. Set $\varepsilon = exp(2\pi i/d)$. We have:

$$A \circ L_1 = L_2 \circ D = L_2 \circ D \circ (\varepsilon z) = A \circ L_1 \circ (\varepsilon z). \tag{33}$$

It follows from Proposition 2.4 that either

$$L_1 \circ (\varepsilon z) = \mu \circ L_1 \tag{34}$$

for some Möbius transformation $\mu(z)$ or, taking into account that $L_1(z)$ is a Laurent polynomial, that

$$L_1 = (az + \frac{b}{z}) \circ z^l, \qquad L_1 \circ (\varepsilon z) = (\tilde{a}z + \frac{\tilde{b}}{z}) \circ z^l \tag{35}$$

for some $a, b, \tilde{a}, \tilde{b} \in \mathbb{C}$ distinct from zero and $l \geq 1$.

In the first case, since $\mu(z)$ transforms infinity to infinity, we see that $\mu(z)$ is a linear function. Furthermore, (34) implies that $\mu^{\circ d} = z$ and therefore $\mu(z) = \omega z$ for some dth root of unity ω. Comparing the coefficients of both parts of formula (34) we see that $L_1(z)$ has the form

$$L_1(z) = z^e L(z^d) \tag{36}$$

for some Laurent polynomial $L(z)$ and $0 \leq e < d$. Set $g = \mathrm{GCD}(e, d)$. It follows from (33), (34) that

$$A \circ (\omega z) = A.$$

Therefore, since $\omega = \varepsilon^e$, the equality

$$A = P \circ z^{d/g} \tag{37}$$

holds for some polynomial $P(z)$.

By (30), (36), (37), we have:

$$A \circ L_1 = P \circ z^{d/g} \circ z^{e/g} L(z^{d/g}) \circ z^g = L_2 \circ z^{d/g} \circ z^g. \tag{38}$$

Therefore,

$$P \circ z^{d/g} \circ z^{e/g} L(z^{d/g}) = L_2 \circ z^{d/g}.$$

Since

$$P \circ z^{d/g} \circ z^{e/g} L(z^{d/g}) = P \circ z^{e/g} L^{d/g} \circ z^{d/g},$$

this implies that

$$L_2 = P \circ z^{e/g} L^{d/g}.$$

Setting now

$$W(z) = z^g, \quad r = e/g, \quad n = d/g$$

we see that equalities (31) hold.

On the other hand, if equalities (35) hold then it follows from

$$A \circ (az + \frac{b}{z}) \circ z^l = L_2 \circ z^d,$$

by Lemma 3.1 that

$$A \circ (az + \frac{b}{z}) = L \circ z^{k/l}, \qquad L_2 = L \circ z^{k/d} \tag{39}$$

for some Laurent polynomial $L(z)$ and $k = \mathrm{LCM}(l, d)$. Let w be a complex number such that $w^2 = b/a$. Since (39) implies that

$$L(w^{k/l} z) \circ z^{k/l} = L((wz)^{k/l}) = A(awz + \frac{b}{wz}) = A(2awz) \circ \frac{1}{2}(z + \frac{1}{z})$$

it follows from Lemma 3.1 that there exists a polynomial $P(z)$ such that

$$A(2awz) = P \circ T_{k/l}, \qquad L(w^{k/l}z) = P \circ \frac{1}{2}(z + \frac{1}{z}). \tag{40}$$

The first equalities in formulas (40), (35) imply respectevely the equalities:

$$A = P \circ T_{k/l} \circ (z/2aw), \qquad L_1 = (2awz) \circ \frac{1}{2}(z + \frac{1}{z}) \circ (z^l/w).$$

Furthermore, by (39), (40)

$$L_2 = P \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\frac{z}{w^{k/l}}) \circ z^{k/d} = P \circ \frac{1}{2}(z + \frac{1}{z}) \circ z^{k/d} \circ (\frac{z}{w^{d/l}}).$$

Since

$$z^d = (w^{d/l}z) \circ z^{k/l} \circ (z^f/w),$$

setting

$$W(z) = z^f/w, \quad n = k/l, \quad m = k/d,$$

where $f = \mathrm{GCD}(l, d)$, we see that equalities (32) hold. $\square$

3.2. **Solutions of equation** (5) **in the case when** $\deg A = \deg B$.

Theorem 3.2. *Suppose that polynomials $A(z), B(z)$ of the same degree and Laurent polynomials $L_1(z), L_2(z)$ satisfy the equation*

$$A \circ L_1 = B \circ L_2. \tag{41}$$

Then either $A \circ L_1 \sim B \circ L_2$ or there exist polynomials $R(z)$, $\tilde{A}(z)$, $\tilde{B}(z)$, $W(z)$ and Laurent polynomials $\tilde{L}_1(z)$, $\tilde{L}_2(z)$ such that

i) $$A = R \circ \tilde{A}, \quad B = R \circ \tilde{B}, \quad L_1 = \tilde{L}_1 \circ W, \quad L_2 = \tilde{L}_2 \circ W,$$

ii) $$\tilde{A} \circ \tilde{L}_1 = \tilde{B} \circ \tilde{L}_2,$$

and

$$\tilde{A} \circ \tilde{L}_1 \sim -T_n \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}), \qquad \tilde{B} \circ \tilde{L}_2 \sim T_n \circ \frac{1}{2}(z + \frac{1}{z}), \tag{42}$$

where $\varepsilon^n = -1$.

Proof. It follows from Proposition 2.4 that either $A \circ L_1 \sim B \circ L_2$ or

$$L_1 = (az + \frac{b}{z}) \circ z^r, \qquad L_2 = (cz + \frac{d}{z}) \circ z^r,$$

for some non-zero $a, b, c, d \in \mathbb{C}$ and $r \geq 1$. Furthermore, in the last case, since $\deg A(z) = \deg B(z)$, the equality (41) implies that $c = \mu_1 a$, $d = \mu_2 b$ for some mth roots of unity μ_1, μ_2, $\mu_1 \neq \mu_2$, where $m = \deg A(z) = \deg B(z)$.

Let w be a solution of the equation $w^2 = b/a$ and ε be a solution of the equation $\varepsilon^2 = \mu_2/\mu_1$. We have:

$$az + \frac{b}{z} = (2awz) \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}) \circ (z/\varepsilon w), \quad cz + \frac{d}{z} = (2aw\varepsilon\mu_1 z) \circ \frac{1}{2}(z + \frac{1}{z}) \circ (z/\varepsilon w).$$

Therefore, without loss of generality we can assume that

$$L_1 = \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}) \circ (z^r/\varepsilon w), \qquad L_2 = \frac{1}{2}(z + \frac{1}{z}) \circ (z^r/\varepsilon w)$$

and concentrate on the equation

$$A \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}) = B \circ \frac{1}{2}(z + \frac{1}{z}). \tag{43}$$

Let $F(z)$ be a Laurent polynomial defined by decomposition (43). Then it follows from

$$F = B \circ \frac{1}{2}(z + \frac{1}{z}) \tag{44}$$

that

$$F(1/z) = F(z).$$

Since

$$F = A \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}) = A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon z)$$

this implies that

$$A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon z) = A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon / z).$$

On the other hand,

$$A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon / z) = A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (z/\varepsilon) = A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon z) \circ (z/\varepsilon^2).$$

Hence, $F(z)$ is invariant with respect to the substitution $z = z/\varepsilon^2$ and therefore

$$F = L \circ z^n, \tag{45}$$

where $L(z)$ is a Laurent polynomial and $n|m$ equals the order of $1/\varepsilon^2$. It follows now from (44), (45) by Lemma 3.1 that

$$B = R \circ T_n, \tag{46}$$

where $R(z)$ is a polynomial.

Substituting (46) in (43) we obtain:

$$A \circ \frac{1}{2}(z + \frac{1}{z}) \circ (\varepsilon z) = R \circ \frac{1}{2}(z^n + \frac{1}{z^n}).$$

Furthermore, substituting in the last equality $z = z/\varepsilon$ and using that $\varepsilon^n = -1$ we conclude that

$$A \circ \frac{1}{2}(z + \frac{1}{z}) = R \circ -\frac{1}{2}(z^n + \frac{1}{z^n}) = R \circ -T_n \circ \frac{1}{2}(z + \frac{1}{z})$$

and, therefore,

$$A = R \circ -T_n. \quad \Box$$

3.3. Reduction of equation (5) for reducible pairs $A(z)$, $B(z)$.

Theorem 3.3. *Suppose that polynomials $A(z), B(z)$ and Laurent polynomials $L_1(z)$, $L_2(z)$ satisfy the equation*

$$A \circ L_1 = B \circ L_2. \tag{47}$$

Then either $A \circ L_1 \sim B \circ L_2$ or there exist polynomials $R(z)$, $\tilde{A}(z)$, $\tilde{B}(z)$, $W(z)$ and Laurent polynomials $\tilde{L}_1(z)$, $\tilde{L}_2(z)$ such that

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

ii) $$\tilde{A} \circ \tilde{L}_1 = \tilde{B} \circ \tilde{L}_2,$$

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

$$\tilde{A} \circ \tilde{L}_1 \sim -T_{nl} \circ \frac{1}{2}(\varepsilon z^m + \frac{\bar{\varepsilon}}{z^m}), \qquad \tilde{B} \circ \tilde{L}_2 \sim T_{ml} \circ \frac{1}{2}(z^n + \frac{1}{z^n}), \tag{48}$$

where $l > 2$, $\mathrm{GCD}(n, m) = 1$, and $\varepsilon^{nl} = -1$.

Proof. Without loss of generality we can assume that the pair $A(z)$, $B(z)$ is reducible and that there exists no rational function $W(z)$, $\deg W(z) > 1$, such that $L_1 = \tilde{L}_1 \circ W$, $L_2 = \tilde{L}_2 \circ W$ for some Laurent polynomials $\tilde{L}_1(z)$, $\tilde{L}_2(z)$.

Furthermore, it follows from Corollary 2.3 that there exist polynomials $A_1(z)$, $B_1(z)$, $\deg A_1 = \deg B_1 > 1$, and $U(z)$, $V(z)$ such that

$$A = A_1 \circ U, \qquad B = B_1 \circ V$$

and the pair $U(z)$, $V(z)$ is irreducible. By Theorem 3.2 applied to the equality

$$A_1 \circ (U \circ L_1) = B_1 \circ (V \circ L_2),$$

either $A_1 \circ (U \circ L_1) \sim B_1 \circ (V \circ L_2)$, and then setting

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

for some $\mu \in \mathrm{Aut}\,(\mathbb{CP}^1)$ we arrive to the first possibility provided by the theorem, or there exist polynomials $R(z)$, $A_2(z)$, $B_2(z)$ such that

$$A_1 = R \circ A_2, \qquad B_1 = R \circ B_2,$$

$$A_2 \circ (U \circ L_1) = B_2 \circ (V \circ L_2),$$

and

$$A_2 \circ (U \circ L_1) \sim -T_l \circ \left(\frac{1}{2}(z + \frac{1}{z}) \circ (\nu c z^d) \right),$$

$$B_2 \circ (V \circ L_2) \sim T_l \circ \left(\frac{1}{2}(z + \frac{1}{z}) \circ (c z^d) \right),$$

where $\nu^l = -1$, $l > 1$, $d \geq 1$, $c \in \mathbb{C}$. Furthermore, passing to appropriate $\tilde{L}_1(z)$, $\tilde{L}_2(z)$ and modifying $A_2(z)$, $B_2(z)$, $U(z)$, $V(z)$ we can assume that

$$U \circ \tilde{L}_1 = \frac{1}{2}(z + \frac{1}{z}) \circ (\nu z^d), \quad V \circ \tilde{L}_2 = \frac{1}{2}(z + \frac{1}{z}) \circ z^d. \tag{49}$$

Applying now Theorem 3.1 to equalities (49) we obtain that

$$U \circ M \sim T_n \circ \frac{1}{2}\left(z^{d/n} + \frac{1}{z^{d/n}} \right), \quad V \circ \tilde{L}_2 \sim T_m \circ \frac{1}{2}\left(z^{d/m} + \frac{1}{z^{d/m}} \right),$$

where $M = \tilde{L}_1 \circ \mu z$ with $\mu^d = 1/\nu$. Therefore,

$$U \circ \tilde{L}_1 \sim T_n \circ \frac{1}{2}\left(\varepsilon z^{d/n} + \frac{\bar{\varepsilon}}{z^{d/n}} \right), \quad V \circ \tilde{L}_2 \sim T_m \circ \frac{1}{2}\left(z^{d/m} + \frac{1}{z^{d/m}} \right),$$

for some $m, n \geq 1$ and $\varepsilon = (1/\mu)^{d/n}$.

Furthermore, in view of the irreducibility of the pair $U(z)$, $V(z)$ the equality $\mathrm{GCD}(n, m) = 1$ holds. Since on the other hand the assumption about $L_1(z)$, $L_2(z)$ implies that $\mathrm{GCD}(d/n, d/m) = 1$ we conclude that $d = nm$ and setting

$$\tilde{A} = A_2 \circ U, \qquad \tilde{B} = B_2 \circ V$$

we obtain that

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

where $\varepsilon^{nl} = -1$.

Finally, $l > 2$ since if $l = 2$ the algebraic curve $T_{2n}(x) + T_{2m}(y) = 0$ is irreducible as one can check using the description of the group $\Gamma_{-T_{2n} \times T_{2m}}$ given in Theorem 2.1. $\square$

3.4. **Solutions of equation** (5) **in the case when the pair** $A(z)$, $B(z)$ **is irreducible.** In this subsection we describe solutions of equation (5) in the case when the pair $A(z), B(z)$ is irreducible.

First of all observe that if $A(z), B(z), C(z), D(z)$ are rational functions satisfying equation $A \circ C = B \circ D$ and the pair $A(z), B(z)$ is irreducible then the genus of the Riemann surface on which the function $h_1(z)$ from Theorem 1 is defined necessarily equals zero. Furthermore, it follows from the construction of $h_1(z)$ that if $A(z), B(z)$ are polynomials then the number of poles of $h_1(z)$ equals $\mathrm{GCD}(\deg A(z), \deg B(z))$. Therefore, the description of the solutions of equation (5) in the case when the pair $A(z), B(z)$ is irreducible essentially is equivalent to the description of all irreducible pairs polynomials $A(z), B(z)$ for which $\mathrm{GCD}(\deg A(z), \deg B(z)) \leq 2$ and the expression

$$g(A,B) = \sum_{j=1}^{o(A,B)} g(R_j)$$

from formula (13) equals 0. Besides, it is necessary to find rational functions $U(z), V(z)$ satisfying $A \circ U = B \circ V$ and such that $\deg U(z) = \deg B(z)$, $\deg V(z) = \deg A(z)$. However, since one can check directly that all the functions $U(z)$, $V(z)$ given below satisfy these requirements we will skip the corresponding calculations.

Theorem 3.4. *Suppose that polynomials $A(z), B(z)$ and Laurent polynomials $L_1(z)$, $L_2(z)$ satisfy equation*

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

and the pair $A(z), B(z)$ is irreducible. Then there exist polynomials $\tilde{A}(z), \tilde{B}(z)$, rational functions $\tilde{L}_1(z)$, $\tilde{L}_2(z)$, $W(z)$, and Möbius transformations $\mu_1(z), \mu_2(z)$ such that

$$A = \mu_1 \circ \tilde{A}, \qquad B = \mu_2 \circ \tilde{B}, \qquad L_1 = \tilde{L}_1 \circ W, \qquad L_2 = \tilde{L}_2 \circ W$$

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

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

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

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

where $T_n(z), T_m(z)$ are Chebyshev polynomials, $m, n \geq 1$, and $\mathrm{GCD}(n, m) = 1$,

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

where $T_{2n_1}(z), T_{2m_1}(z)$ are Chebyshev polynomials, $m_1, n_1 \geq 1$, $\varepsilon^{2n_1} = -1$, and $\mathrm{GCD}(n_1, m_1) = 1$,

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

where $S(z)$ is a polynomial,

5) $$\tilde{A} \circ \tilde{L}_1 \sim (z^2-1)^3 \circ \frac{3(3z^4+4z^3-6z^2+4z-1)}{(3z^2-1)^2},$$

$$\tilde{B} \circ \tilde{L}_2 \sim (3z^4-4z^3) \circ \frac{4(9z^6-9z^4+18z^3-15z^2+6z-1)}{(3z^2-1)^3}.$$

The proof of this theorem is given below and consists of the following stages. First we rewrite formula for the genus in a more convenient way and prove several lemmas. After introducing the conception of a special point the rest of the proof splits into two parts: when only one of polynomials $A(z), B(z)$ has a special point and when both $A(z), B(z)$ have a special point.

3.4.1. *Genus formula and lemmas.* Working with polynomials it is convenient to "forget" about infinite critical values in the following sense: if $B(z)$ is a polynomial and $z_1, z_2, \dots, z_r$ is a set of all critical values of $B(z)$ then we will associate to $B(z)$ a collection of partitions

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

of $m = \deg B(z)$ corresponding to the decompositions of the permutations $\alpha_i(B)$, $1 \leq i \leq r$, into products of disjoint cycles for the finite points $z_1, z_2, \dots, z_s$, $s = r-1$, only. We will call such a collection the passport of $B(z)$.

In the following the set $z_1, z_2, \dots, z_s$ will denote a union of all finite critical values of a pair polynomials $A(z), B(z)$. Therefore, some of partitions above may contain only units. We will call such partitions trivial and will denote by $s(B)$ the number of non-trivial partitions. Clearly, by the Riemann-Hurwitz formula we have:

$$\sum_{i=1}^{s} q_i = (s-1)m + 1. \tag{50}$$

Lemma 3.2. *Let*

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

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

be passports of polynomials $A(z), B(z)$, $n = \deg A(z)$, $m = \deg B(z)$. Then

$$-2g(A,B) = \mathrm{GCD}(m,n) - 1+$$

$$+\sum_{i=1}^{s}\sum_{j_1=1}^{p_i}\left[a_{i,j_1}(1-q_i) - 1 + \sum_{j_2=1}^{q_i}\mathrm{GCD}(a_{i,j_1}b_{i,j_2})\right]. \tag{51}$$

Proof. Indeed, we have:

$$\sum_{i=1}^{s}\sum_{j_1=1}^{p_i}[a_{i,j_1}(1-q_i)-1] = \sum_{i=1}^{s}[n(1-q_i)-p_i] = ns - n\sum_{i=1}^{s}q_i - \sum_{i=1}^{s}p_i =$$

$$= ns - n((s-1)m+1) - ((s-1)n+1) = -n(s-1)m - 1.$$

Therefore, the right side of formula (51) 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 (51) follows from (13) taking into account that $r = s+1$. $\square$

Set

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

$1 \le i \le s$, $1 \le j_1 \le p_i$. In this notation formula (51) takes the form

$$-2g(A,B) = \mathrm{GCD}(m,n) - 1 + \sum_{i=1}^{s}\sum_{j_1=1}^{p_i} s_{i,j_1}. \tag{52}$$

Lemma 3.3. *In the notation of Lemma 3.2 for any i, j_1, $1 \le i \le s$, $1 \le j_1 \le p_i$, the following statements hold:*

a) If there exist at least three numbers $b_{i,l_1}, b_{i,l_2}, b_{i,l_3}$, $1 \le l_1, l_2, l_3 \le q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1} \le -2$.

b) If there exist exactly two numbers b_{i,l_1}, b_{i,l_2}, $1 \le l_1, l_2 \le q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1} \le -1$ and the equality attains if and only if

$$\mathrm{GCD}(a_{i,j_1} b_{i,l_1}) = \mathrm{GCD}(a_{i,j_1} b_{i,l_2}) = a_{i,j_1}/2, \tag{53}$$

c) If there exists exactly one number b_{i,l_1} which is not divisible by a_{i,j_1} then

$$s_{i,j_1} = -1 + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}). \tag{54}$$

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

$$s_{i,j_1} = a_{i,j_1}(1-q_i) - 1 + \sum_{\substack{j_2=1\\ j_2 \ne l_1,l_2,l_3}}^{q_i} \mathrm{GCD}(a_{i,j_1} b_{i,j_2}) + \sum_{l_1,l_2,l_3} \mathrm{GCD}(a_{i,j_1} b_{i,l_1}) \le$$

$$\le a_{i,j_1}(1-q_i) - 1 + (q_i - 3)a_{i,j_1} + 3a_{i,j_1}/2 = -a_{i,j_1}/2 - 1 < -1.$$

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

$$s_{i,j_1} = a_{i,j_1}(1-q_i) - 1 + \sum_{\substack{j_2=1\\ j_2 \ne l_1,l_2}}^{q_i} \mathrm{GCD}(a_{i,j_1} b_{i,j_2}) + \sum_{l_1,l_2} \mathrm{GCD}(a_{i,j_1} b_{i,l_1}) \le$$

$$\le a_{i,j_1}(1-q_i) - 1 + (q_i - 2)a_{i,j_1} + a_{i,j_1}/2 + a_{i,j_1}/2 = -1,$$

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:

$$s_{i,j_1} = a_{i,j_1}(1-q_i) - 1 + \sum_{\substack{j_2=1\\ j_2 \ne l_1}}^{q_i} \mathrm{GCD}(a_{i,j_1} b_{i,j_2}) + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}) =$$

$$= -1 + \mathrm{GCD}(a_{i,j_1} b_{i,l_1}). \quad \square$$

Lemma 3.4. *In the same notation suppose that*

$$\mathrm{GCD}(b_{i1}, b_{i2}, ..., b_{iq_i}) = 1 \tag{55}$$

for any i, $1 \le i \le s$. Then for any i, j_1, $1 \le i \le s$, $1 \le j_1 \le p_i$, we have:

a) $s_{i,j_1} \le 0$,

b) $s_{i,j_1} = 0$ if and only if either $a_{i,j_1} = 1$ or all the numbers b_{i,j_2}, $1 \le j_2 \le q_i$, 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}, $1 \le j_2 \le q_i$, *except two are even.*

Proof. If $a_{i,j_1} = 1$ then $s_{i,j_1} = 0$ so assume that $a_{i,j_1} > 1$. If there exists exactly one number b_{i,l_1} which is not divisible by a_{i,j_1} then in view of (55) necessarily $\text{GCD}(a_{i,j_1} b_{i,l_1}) = 1$ and hence $s_{i,j_1} = 0$ by (54).

If there exist exactly two numbers b_{i,l_1}, b_{i,l_2}, $1 \le l_1, l_2 \le q_i$, which are not divisible by a_{i,j_1} then by Lemma 3.3 $s_{i,j_1} \le -1$ where the equality attains if and only if (53) holds. This implies that $a_{i,j_1} = 2$ since otherwise we obtain a contradiction with (55).

Finally, if there exist at least three numbers $b_{i,l_1}, b_{i,l_2}, b_{i,l_3}$, $1 \le l_1, l_2, l_3 \le q_i$, which are not divisible by a_{i,j_1} then $s_{i,j_1} \le -2$ by Lemma 3.3. $\square$

Say that a polynomial $B(z)$ have a special point if there exists i, $1 \le i \le s$, such that

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

Say that a polynomial $B(z)$ have a 1-special point (resp. a 2-special point) if there exists i, $1 \le i \le s$, such that all the numbers

$$b_{i,1}, b_{i,2}, ..., b_{i,q_i}$$

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

Proposition 3.1. *Let $B(z)$ be a polynomial. Then*

a) $B(z)$ may not have two special points, or one special point and one 1-special point, or more than two 1*-special points,*

b) if $B(z)$ has two 1*-special points then $s(B) = 2$ and the corresponding non-trivial partitions are* $(1, 2, \ldots 2)$, $(1, 2, \ldots 2)$,

c) if $B(z)$ has one 1-special point and one 2-special point then $s(B) = 2$ and the corresponding non-trivial partitions are either $(1, 1, 2)$, $(1, 3)$ *or* $(1, 2, 2)$, $(1, 1, 3)$.

Proof. Let

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

be a collection of partitions corresponding to $B(z)$, $m = \deg B(z)$. Suppose first that $B(z)$ has at least two 1-special points. To be definite assume that the corresponding indices are $1, 2$ and that all $(b_{1,1}, \ldots, b_{1,q_1})$ but $b_{1,1}$ are divisible by the number d_1 and all $(b_{2,1}, \ldots, b_{2,q_2})$ but $b_{2,1}$ are divisible by the number d_2. Then

$$q_1 \le 1 + \frac{m - b_{1,1}}{d_1}, \qquad q_2 \le 1 + \frac{m - b_{2,1}}{d_2}, \tag{56}$$

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

$$\sum_{i=1}^{s} q_i \le q_1 + q_2 + (s-2)m, \tag{57}$$

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

$$q_i \le 1 + \frac{m - b_{i,1}}{d_i} \le 1 + \frac{m-1}{2} \tag{58}$$

and hence

$$q_1 + q_2 \leq 1 + m \tag{59}$$

where the equality attains only if $d_1 = 2$, $d_2 = 2$, $b_{1,1} = 1$, $b_{2,1} = 1$. Now (57) and (59) imply that

$$\sum_{i=1}^{s} q_i \leq (s-1)m + 1. \tag{60}$$

Since however in view of (50) in this inequality should attain equality we conclude that in all intermediate inequalities should attain equalities and therefore $s(B) = 2$ and

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

In particluar, we see that $B(z)$ may not have more than two 1-special points.

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

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

Since this inequality is stronger than (58) repeating the argument above we obtain an inequality in (60) in contradiction with (50).

Finally, suppose that the index 1 corresponds to a 1-special point while the index 2 corresponds to a 2-special point. We will suppose that 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 odd then $d_2 > 2$ and we have:

$$q_1 \leq 1 + \frac{m - b_{1,1}}{d_1} \leq 1 + \frac{m-1}{2}, \quad q_2 \leq 2 + \frac{m - b_{2,1} - b_{2,2}}{d_2} \leq 2 + \frac{m-2}{3}.$$

Therefore,

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

If $m > 5$ then

$$\frac{11}{6} + \frac{5m}{6} < m + 1$$

that together with (57) gives a contradiction with (50).

On the other hand, if $m \leq 5$ then one can check directly that a unique possibility for $B(z)$ to have one 1-special point and one 2-special point is the one corresponding to the passport

$$(b_{1,1}, \dots, b_{1,q_1}) = (1, 2, 2), \quad (b_{2,1}, \dots, b_{2,q_2}) = (1, 1, 3).$$

Similarly, if m is even then $d_1 > 2$ and we have:

$$q_1 \leq 1 + \frac{m - b_{1,1}}{d_1} \leq 1 + \frac{m-1}{3}, \quad q_2 \leq 2 + \frac{m - b_{2,1} - b_{2,2}}{d_2} \leq 2 + \frac{m-2}{2}.$$

Therefore,

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

If $m > 4$ then

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

and as above we obtain a contradiction with (50). On the other hand, one can check that if $m \leq 4$ then we should have:

$$(b_{1,1},\dots,b_{1,q_1}) = (1,1,2), \qquad (b_{2,1},\dots,b_{2,q_2}) = (1,3). \quad \square$$

3.4.2. *Proof of Theorem 3.4. Part 1.* In this subsection we prove Theorem 3.4 under the assumption that at least one of polynomials $A(z), B(z)$ does not have special points. Without loss of generality we can suppose that this polynomial is $B(z)$.

Suppose first that $\mathrm{GCD}(n,m) = 1$. In this case by formula (52) the condition $g(A,B) = 0$ is equivalent to the condition

$$\sum_{i=1}^{s}\sum_{j_1=1}^{p_i} s_{i,j_1} = 0. \tag{61}$$

In view of Lemma 3.4,a this is possible if and only of $s_{i,j_1} = 0$, $1 \leq i \leq s$, $1 \leq j_1 \leq p_i$.

If $A(z)$ has a unique finite critical value then Lemma 3.4,b implies that we should have 1). Furthermore, it follows from Lemma 3.4,b and Proposition 3.1,a,b that if the polynomial $A(z)$ has at least two finite critical values then actually $A(z)$ has exactly two critical values, $B(z) = T_m(z)$ and

$$a_{1,j_1} \leq 2, \quad a_{2,j_2} \leq 2, \quad 1 \leq j_1 \leq p_1, \quad 1 \leq j_2 \leq p_2. \tag{62}$$

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 easily from (62) that the corresponding partitions for $A(z)$ are either $(1,2,2,...,2)$, $(1,2,2,...,2)$ or $(1,1,2,...,2)$, $(2,2,2,...,2)$. Therefore, taking into account a well know fact that for any polynomial $P(z)$ with such a passport there exist Möbius transformations $\mu_1(z), \mu_2(z)$ such that $\mu_1 \circ P \circ \mu_2 = T_n$, we see that in this case we arrive to 2).

Suppose now that $\mathrm{GCD}(n,m) = 2$. Then the condition $g(A,B) = 0$ is equivalent to the condition that one number from s_{i,j_1}, $1 \leq i \leq s$, $1 \leq j_1 \leq p_i$, equals -1 while others equal 0. If $A(z)$ has one critical value than it follows easily from Lemma 3.4,c that we should have 4).

On the other hand, if $A(z)$ has at least two critical values then Lemma 3.4, b,c and Proposition 3.1,a,c, taking into account that $\deg B(z)$ is even in view of $\mathrm{GCD}(n,m) = 2$, imply that the partitions corresponding to $B(z)$ are $(1,3)$, $(1,1,2)$. Furthermore, we see that for any j_1, $1 \leq j_1 \leq p_1$, the number a_{1,j_1} equals 1 or 3 and that 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 corresponding to $A(z)$ and by γ the number of appearance of 1 in the second partition. We have:

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

and, by (50)

$$\alpha + \beta + \gamma = n.$$

The second and the third equations imply that $\alpha+\beta=2$. This implies easily that the partitions corresponding to $A(z)$ are either $(1,3)$, $(1,1,2)$ or $(3,3)$, $(2,1,1,1,1)$.

It is not hard to prove however that for any polynomial $R(z)$ with the partitions $(1,3)$, $(1,1,2)$ there exist Möbius transformations $\mu_1(z),\mu_2(z)$ such that $\mu_1\circ R\circ\mu_2=3z^4-4z^3$. Therefore, the first case is not possible since otherwise $A(z)=B(z)$ and the curve $A(x)-B(y)=0$ is reducible. On the other hand, it is easy to check that in the second case we should have $\mu_1\circ A\circ\mu_2=(z^2-1)^3$.

3.4.3. *Proof of Theorem 3.4. Part 2.* Suppose now that both polynomials $A(z)$ and $B(z)$ have special points. Then by Proposition 3.1 each of them has a unique special point. The special points of $A(z)$ and $B(z)$ either coincide or are different. In the second case without loss of generality we can assume that

$$A=(z^{d_1}+\beta_1)\circ U,\qquad B=(z^{d_2}+\beta_2)\circ V, \tag{63}$$

for some $\beta_1,\beta_2\in\mathbb{C}$, $\beta_1\neq\beta_2$, and $d_1,d_2>1$. Since the pair $A(z)$, $B(z)$ is irreducible and $g(A,B)=0$ the same is true for the pair $A_1(z)=z^{d_1}+\beta_1$, $B_1(z)=z^{d_2}+\beta_2$ and hence

$$g(A_1,B_1)=0. \tag{64}$$

Formula (51) implies that

$$2-2g(A_1,B_1)=d_1+d_2-d_1d_2+\mathrm{GCD}(d_1,d_2). \tag{65}$$

If $\mathrm{GCD}(d_1,d_2)=1$ then (64) 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 (64) is equivalent to the equality $(d_1-1)(1-d_2)=-1$ which holds if and only if $d_1=d_2=2$.

Since

$$A_1\circ U\circ L_1=B_1\circ V\circ L_2$$

and $\deg A_1=\deg B_1$, using now the same reasoning as in the proof of Theorem 3.3 and taking into account the condition $\mathrm{GCD}(d_1,d_2)=2$ we arrive to 3).

In the case when the special points of $A(z)$ and $B(z)$ coincide we can assume without loss of generality that

$$A=z^{d_1}\circ U,\qquad B=z^{d_2}\circ V, \tag{66}$$

where

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

and

$$\mathrm{GCD}(d_1,d_2)=1 \tag{67}$$

in view of the irreducibility of the pair $A(z)$ and $B(z)$. Without loss of generality we may assume that

$$\deg U(z)>1,\quad \deg V(z)>1 \tag{68}$$

since otherwise Lemma 3.3 implies easily that we should have 1) or 4). Finally, without loss of generality we can assume that $m=\deg B(z)$ is greater than $n=\deg A(z)$. 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 $A(z),B(z)$ with $g(A,B)=0$.

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

$$x^{d_1} - B(y) = 0 \tag{69}$$

is an irreducible curve of genus zero, formula (52) and Lemma 3.3 applied to polynomials z^{d_1} and $B(z)$, taking into account (67), imply that all the numbers $b_{1,1}, b_{1,2}, ..., b_{1,q_1}$ but two, say b_{1,q_1-1}, b_{1,q_1}, are divisible by d_1 and

$$\mathrm{GCD}(d_1, b_{1,q_1-1}) = \mathrm{GCD}(d_1, b_{1,q_1}) = 1. \tag{70}$$

Returning to polynomials $A(z)$, $B(z)$ we see that, since each a_{1,j_1}, $1 \leq j_1 \leq p_1$, is divisible by d_1, equality (70) implies that

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

$$\leq -a_{1,j_1} - 1 + \mathrm{GCD}(a_{1,j_1} b_{1,q_1-1}) + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \leq$$

$$\leq -a_{1,j_1} - 1 + a_{1,j_1}/d_1 + a_{1,j_1}/d_1 \leq -a_{1,j_1} - 1 + a_{1,j_1}/2 + a_{1,j_1}/2 \leq -1.$$

Since by assumption $p_1 \geq 2$ and by lemma 3.4

$$\sum_{j_1=1}^{p_i} s_{i,j_1} \leq 0, \quad 1 < i \leq s, \quad 1 \leq j_1 \leq p_i,$$

it follows now from formula (52) that $g(A, B) < 0$.

Case 2. If $\mathrm{GCD}(d_1, m) = 1$ then applying as above formula (52) and Lemma 3.3 to polynomials z^{d_1} and $B(z)$ we conclude that each b_{1,j_1}, $1 \leq j_1 \leq q_1$, except one, say b_{1,q_1}, is divisible by d_1 and $\mathrm{GCD}(b_{1,q_1}, d_1) = 1$.

Since each a_{1,j_1}, $1 \leq j_1 \leq p_1$, is divisible by d_1, this implies that

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

$$\leq -1 + \mathrm{GCD}(a_{1,j_1} b_{1,q_1}) \leq -1 + a_{1,j_1}/d_1, \tag{71}$$

and hence

$$\sum_{j_1=1}^{p_1} s_{1,j_1} \leq -p_1 + n/d_1. \tag{72}$$

Furthermore, since each b_{1,j_2}, $1 \leq j_2 \leq q_1$, is divisible by d_2 and each b_{1,j_2}, $1 \leq j_2 \leq q_1$, except one is divisible by d_1 we have:

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

and therefore

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

Since by (50) the inequality

$$q_1 + q_i \geq m + 1 \tag{73}$$

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

$$q_i \geq m - m/d_1 d_2 + 1/d_1. \tag{74}$$

Denote by γ_i the number of units among the numbers b_{i,j_2}, $1 \leq j_2 \leq q_i$. Since the number of non units is $\leq m/2$ the equality $\gamma_i \geq q_i - m/2$ holds and therefore (74) implies that

$$\gamma_i \geq m/2 - m/d_1d_2 + 1/d_1. \tag{75}$$

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

$$s_{i,j_1} \leq a_{i,j_1}(1-q_i) - 1 + a_{i,j_1}(q_i - \gamma_i) + \gamma_i = (1-\gamma_i)(a_{i,j_1} - 1). \tag{76}$$

Hence,

$$\sum_{j_1=1}^{p_i} s_{i,j_1} \leq (1-\gamma_i)(n-p_i) \leq (1 - 1/d_1 + m(1/d_1d_2 - 1/2))(n-p_i)$$

in view of (75). Therefore, using (50) we obtain

$$\sum_{i=2}^{s}\sum_{j_1=1}^{p_i} s_{i,j_1} \leq (1 - 1/d_1 + m(1/d_1d_2 - 1/2))(p_1 - 1). \tag{77}$$

Set

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

By formula (52), in order to finish the proof it is enough to show that $S < -1$.

It follows from (72), (77) that

$$S \leq -p_1 + n/d_1 + (1 - 1/d_1 + m(1/d_1d_2 - 1/2))(p_1 - 1) <$$

$$< -1 + n/d_1 + m(1/d_1d_2 - 1/2)(p_1 - 1). \tag{78}$$

If $p_1 \geq 3$ then (78), taking into account the assumption $m \geq n$, implies that

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

If $p_1 = 2$ then (78) implies that

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

and if $d_1 > 2$ we obtain again that $S < -1$. Finally, if $p_1 = 2$, $d_1 = 2$ but $m \geq (3/2)n$ then similarly (78) implies

$$S < -1 + n(3/4d_2 - 1/4) \leq -1,$$

since in this case $d_2 \geq 3$ by (67).

Therefore, the only case when the proof is still not finished is the one when $p_1 = 2$, $d_1 = 2$, and $n \leq m < (3/2)n$. In this case switch $A(z)$ and $B(z)$ keeping the same notation. This means that we should consider the case when $q_1 = 2$, $d_2 = 2$ and

$$2n/3 < m \leq n. \tag{79}$$

In this case (73) implies that $q_i \geq m - 1$. Therefore, the corresponding partition of m is either trivial or has the form $(1, 1, \ldots, 1, 2)$ and hence $\gamma_i \geq m - 2$. It follows now from (76) that

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

Since in view of the condition $d_2 = 2$ the inequality $d_1 \geq 3$ holds, this implies that

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

If $n \geq 6$ then this inequality implies that $S < -1$. On the other hand, if $n \leq 5$ then, taking into account inequality (79), it is easy to see that there exist no polynomials $A(z)$, $B(z)$ satisfying (66), (68) and $\mathrm{GCD}(n, m) = 1, 2$. $\square$

3.5. **Proof of Theorem 1.1.** Since the use of the Möbius transformations reduces the problem of description of double decompositions of functions from H to the similar problem for Laurent polynomial and any double decomposition of a Laurent polynomial is equivalent to (4), (5) or (6), the first part of Theorem 1.1 follows from Theorems 3.1, 3.3, 3.4 and Lemma 3.1. The second part follows from the proposition below.

Proposition 3.2. *The class $\mathcal{R}_2$ is the Ritt class.*

Proof. The classification of double decompositions of functions from $\mathcal{R}_2$ implies that any elementary double decomposition $A \circ C = B \circ D$ contained in $\mathcal{R}_2$ and such that the pair $A(z), B(z)$ is reducible is related via Möbius transformations to the decomposition

$$-T_l \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}) = T_l \circ \frac{1}{2}(z + \frac{1}{z}),$$

where $\varepsilon^l = -1$.

It follows now from Theorems 2.5 that in order to prove the proposition it is enough to check that for any choice of maximal decompositions

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

the decompositions

$$-u_d \circ u_{d-1} \circ \cdots \circ u_1 \circ \frac{1}{2}(\varepsilon z + \frac{\bar{\varepsilon}}{z}), \qquad v_l \circ v_{l-1} \circ \cdots \circ v_1 \circ \frac{1}{2}(z + \frac{1}{z}) \tag{80}$$

are weakly equivalent.

It is not hard to prove 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 $d_1 d_2 ... d_s = l$. This implies easily that both decompositions (80) are weakly equivalent to some decomposition of the form

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

References

1. R. Avanzi, U. Zannier, *The equation $f(X) = f(Y)$ in rational functions $X = X(t)$, $Y = Y(t)$*, Compos. Math. 139, No. 3, 263-295 (2003)
2. Y. Bilu, R. Tichy, *The Diophantine equation $f(x) = g(y)$*, Acta Arith. 95, No.3, 261-288 (2000).
3. M. Briskin, N. Roytvarf, Y. Yomdin, *Center conditions at infinity for Abel differential equation*, preprint.
4. J. M. Couveignes, L. Granboulan, *Dessins from a geometric point of view*, in The Grothendieck theory of dessins d'enfants, 79-113. Cambridge University Press, 1994.
5. J. Gutierrez, D. Sevilla, *Building counterexamples to generalizations for rational functions of Ritt's decomposition theorem*, J. Algebra 303, No. 2, 655-667 (2006).
6. M. Fried, *On a theorem of Ritt and related diophantine problems*, J. Reine Angew. Math. 264, 40-55 (1973).

7. M. Fried, *Fields of definition of function fields and a problem in the reducibility of polynomials in two variables*, Commun. Algebra 5, 17-82 (1977).
8. F. Klein, *Lectures on the icosahedron and the solution of equations of the fifth degree*, New York: Dover Publications, (1956).
9. A. Kurosch, *The theory of groups. Vol. I,* New York: Chelsea Publishing Company, (1955).
10. N. Magot, A. Zvonkin, *Belyi functions for Archimedean solids*, Discrete Math. 217, No.1-3, 249-271 (2000).
11. W. Massey, *Algebraic topology: An introduction*, Graduate Texts in Mathematics, Vol. 56, (1981).
12. R. Miranda, *Algebraic curves and Riemann surfaces*, Graduate Studies in Mathematics. 5. Providence, RI: AMS, American Mathematical Society, (1995).
13. M. Muzychuk, F. Pakovich, *Solution of the polynomial moment problem*, preprint, arXiv:0710.4085.
14. F. Pakovich, *On the functional equation $F(A(z)) = G(B(z))$, where A,B are polynomial and F, G are continuous functions*, Math. Proc. Cam. Phil. Soc., 143, No. 2, 469-472 (2007).
15. J Ritt, *Prime and composite polynomials*, American M. S. Trans. 23, 51-66 (1922).
16. J Ritt, *Equivalent rational substitutions*, American M. S. Trans. 26, 221-229 (1924).
17. A. Schinzel, *Polynomials with special regard to reducibility*, Encyclopedia of Mathematics and Its Applications **77**, Cambridge University Press, 2000.
18. U. Zannier, *Ritt's second theorem in arbitrary characteristic,* J. Reine Angew. Math. 445, 175-203 (1993)
19. M. Zieve, *Private communication.*

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

E-mail address: `pakovich@math.bgu.ac.il`