Ritt's theory on the unit disk

By Tuen Wai Ng* and Ming-Xi Wang**

*Department of Mathematics,

The University of Hong Kong, Pokfulam, Hong Kong.

E-mail address: ntw@maths.hku.hk

**Department of Mathematics,

ETH, Zürich.

E-mail address: wangmx@math.ethz.ch

Abstract

The aim of this paper is to revisit Ritt's theory [35] from a topological perspective by extensively using the concept of fundamental groups. This enables us to regard the theory as an example which illustrates many ideas of Grothendieck's letter [20] and to put Ritt's theory into a more general analytic setting. In particular, Ritt's theory on the unit disk will be carefully developed and a special class of finite Blaschke products will be introduced as the counterpart of Chebyshev polynomials in Ritt's theory. These finite Blaschke products will be shown to be closely related to the elliptic rational functions which are of great importance in the filter design theory.

1 Introduction

We call a nonlinear polynomial f in $\mathbb{C}[z]$ *prime* if there do not exist nonlinear polynomials φ_1 and φ_2 in $\mathbb{C}[z]$ for which $f = \varphi_1 \circ \varphi_2$. Otherwise f is called *composite* or *factorized*. A representation of f in the form $f = \varphi_1 \circ \cdots \circ \varphi_k$ is a *factorization* or *decomposition* of f and a maximal factorization of f into prime polynomials only is called a *prime factorization* of f. The *length* of f, with respect to a given prime factorization, is defined to be the number of prime polynomials present in that prime factorization. In 1922 J.F. Ritt [35] proved three fundamental results on factorizations of complex polynomials.

AMS Classification: Primary, 30D05; Secondary, 39B12,30J10.

Key words and phrases: Ritt's theory, finite maps, fundamental groups, monodromy, finite Blaschke products, Jacobian elliptic functions.

* Partially supported by the RGC grant HKU 7020/03P and HKU 704409P.

** Partially supported by OYR award of HKU, RGC grant HKU 7020/03P and HKU 704409P, graduate studentship of HKU and Schweizer Nationalfonds zur Förderung der Wissenschaften Grant PDFMP2_122850/1..

He first gives a necessary and sufficient condition for a complex polynomial to be composite and shows that *a nonlinear polynomial f in $\mathbb{C}[z]$ is composite if and only if its monodromy group is imprimitive* (Ritt I), and that *the length of a nonlinear polynomial f in $\mathbb{C}[z]$ is independent of its prime factorizations* (Ritt II). The third result of Ritt tells us how to pass from one prime factorization to another one. Here we use $\mathrm{Aut}_{\mathbb{C}}(\mathbb{C})$ to denote the set of complex polynomials of degree 1 and T_k to denote the Chebyshev polynomial of degree k.

Ritt's Theorem (Ritt III). *Given two prime factorizations of a nonlinear polynomial $f \in \mathbb{C}[z]$, one can pass from one prime factorization to the other one by repeatedly uses of the following operations:*

1) $h \circ g = (h \circ \iota^{-1}) \circ (\iota \circ g)$ *with* $h, g \notin \mathrm{Aut}_{\mathbb{C}}(\mathbb{C})$ *non-constant polynomials and* ι *an element in* $\mathrm{Aut}_{\mathbb{C}}(\mathbb{C})$*;*

2) $T_p \circ T_q = T_q \circ T_p$ *with* p, q *prime numbers;*

3) $z^r g(z)^k \circ z^k = z^k \circ z^r g(z^k)$ *with* r, k *in* $\mathbb{N}$ *and* g *a non-constant polynomial.*

After Ritt's original work, many authors have tried to give different proofs or generalizations of Ritt's theorems. H.T. Engstrom [13] and H. Levi [23] proved (Ritt I) in 1941 and (Ritt II) in 1942 respectively in the case of an arbitrary field of characteristic zero and in 1974 F. Dorey and G. Whaples [12] reformulated the work by adopting the valuation theory. U. Zannier [40] settled the case of fields of positive characteristics in 1993 and P.F. Müller gave in [26] a group-theoretic proof of the Ritt theorems in 1995. A new factorization invariant was discovered by A.F. Beardon and T.W. Ng [5] in 2000 and recently F. Pakovich [30] obtained very interesting results on the factorization of Laurent polynomials. There is a lattice structure hidden in the problem and this was extensively studied by M. Muzychuk and F. Pakovich [28]. A very subtle relationship between decompositions and iterations was revealed in a joint paper by M. Zieve and P. Müller [27].

On the one hand all these studies are based on algebraic techniques and on the other hand Ritt's original work is simply topological in nature. We shall adopt Ritt's topological point of view and explore the theory by means of topological fundamental groups. This enable us to put Ritt's theory in a more general analytic setting and the main goal of this paper is to develop a version of Ritt's theory for the unit disk.

Ritt's theory is closely related to questions about rational points on curves. This was first observed by M. Fried [14] who applied Ritt's theory in arithmetics to study integral points on curves, for the case of rational points see Avanzi-Zannier

[4]. The work of M. Fried was completed by Yu.F. Bilu and R.F. Tichy in [6] and a remarkable application of the theory combined with the Bilu-Tichy Criterion to arithmetic dynamics can be found in a recent paper by D. Ghioca, T.J. Tucker and M.E. Zieve in [18]. New applications of Ritt's theory in function theory are published in Dinh's paper [11] on sharing sets and in Pakovich's work [29].

In analytic geometry or algebraic geometry a *finite map* refers to a an analytic or algebraic map which is proper and quasi-finite. As a special case a holomorphic map between Riemann surfaces is finite if and only if it is non-constant and proper. This notion was first introduced by Radó who proved in [31] that *a holomorphic map* $f : \mathfrak{M} \to \mathfrak{N}$ *between Riemann surfaces is finite if and only if there exists an integer* n *such that* $f(z) = c$ *has* n *solutions for all* $c \in \mathfrak{N}$ and we refer the reader to [17, p.27] for a modern treatment. We shall define the number n given above to be the degree of f and denote it by $\deg f$. One may deduce readily from [8, p.99] that *if* $h : \mathfrak{M} \to \mathfrak{T}$ *and* $g : \mathfrak{T} \to \mathfrak{N}$ *are holomorphic maps between Riemann surfaces, then* $g \circ h$ *is finite if and only if both* g *and* h *are finite.* This suggests that it is natural to study factorizations of finite maps. Since non-constant polynomials are all finite self-maps of the complex plane, Ritt's original theory fits into this more general setting.

A finite map is called *linear* if $\deg f = 1$. If f is a nonlinear finite map from $\mathfrak{M}$ to $\mathfrak{N}$ be a nonlinear finite map then we call f *prime* if there do not exist nonlinear finite maps $\varphi_1 : \mathfrak{T} \to \mathfrak{N}$ and $\varphi_2 : \mathfrak{M} \to \mathfrak{T}$ for which $f = \varphi_1 \circ \varphi_2$. Otherwise it is called *composite* or *factorized.* We shall call a factorization of f *proper* if all its factors are nonlinear and a maximal proper factorization is called a *prime factorization.* The *length* of f with respect to a prime factorization is defined to be the number of its factors. Then Ritt's first two theorems can be reformulated for finite maps.

Theorem 1.1 (Ritt I$'$)**.** *If* f *is a nonlinear finite map from* $\mathfrak{M}$ *to* $\mathfrak{N}$ *then it is composite if and only if its monodromy group is imprimitive.*

For our version of (Ritt II) we need an additional hypothesis, which is satisfied for all finite maps with a totally ramified point, in particular for polynomial maps.

Theorem 1.2 (Ritt II$'$)**.** *If* $\alpha : [0, 1] \to \mathfrak{N}$ *is a closed cycle on* $\mathfrak{N}$ *over which* f *is unramified and if the monodromy of* α *acts transitively then the length of* f *is independent of the prime factorizations.*

The proofs are only slight technical modifications of the original proofs to deal with the more general situation. We shall apply these two theorems when the Riemann surfaces $\mathfrak{M}$ and $\mathfrak{N}$ are unit disks $\mathbb{E}$ and carefully develop a complete version of Ritt's theory on $\mathbb{E}$. Since Chebyshev polynomials play an important role in Ritt's theory, it is natural to find their counterparts in the unit disk case. We solve this central problem by introducing in Section 5 a new class of finite Blaschke products, which we call Chebyshev-Blaschke products $f_{n,t}$ for $n \in \mathbb{N}$ and $t > 0$.

Main Theorem 1.3. *Let f be a finite map from $\mathbb{E}$ to $\mathbb{E}$,*

$$\mathbb{E} \overset{\varphi_1}{\to} \mathfrak{T}_1 \overset{\varphi_2}{\to} \mathfrak{T}_2 \to \quad \cdots \quad \to \mathfrak{T}_{r-1} \overset{\varphi_r}{\to} \mathbb{E}$$

and

$$\mathbb{E} \overset{\psi_1}{\to} \mathfrak{V}_1 \overset{\psi_2}{\to} \mathfrak{V}_2 \to \quad \cdots \quad \to \mathfrak{V}_{s-1} \overset{\psi_s}{\to} \mathbb{E}$$

decompositions of f into a product of prime finite maps. We can pass from the first decomposition to the second by applying repeatedly the following operations:

1) $h \circ g = (h \circ \iota^{-1}) \circ (\iota \circ g)$ *with h, g nonlinear finite endomorphisms of $\mathbb{E}$ and ι a linear map from $\mathbb{E}$ to another Riemann surface;*

2) $(\iota \circ f_{p,qt}) \circ (f_{q,t} \circ \jmath) = (\iota \circ f_{q,pt}) \circ (f_{p,t} \circ \jmath)$ *with p, q prime numbers, t a positive real number and $\iota, \jmath$ elements in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$;*

3) $(\iota \circ z^r g(z)^k) \circ (z^k \circ \jmath) = (\iota \circ z^k) \circ (z^r g(z^k) \circ \jmath)$, *with r, k rational integers, g a finite endomorphism of $\mathbb{E}$ and $\iota, \jmath$ elements in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$.*

In Section 2 we introduce some basic results on finite maps and their monodromy groups. We then modify Ritt's proof of his first two theorems to deal with the case of finite maps (Theorem 1.1 and Theorem 1.2) in Section 3. In Section 4 we show how to deform a finite map between unit disks to obtain a polynomial. By making use of monodromy groups we introduce in Section 5 the Chebyshev-Blaschke products $f_{n,t}$ and we shall then explain carefully in Section 6 and 7 how $f_{n,t}$ can be expressed in terms of elliptic rational functions which are intensively used in filter design theory. The reader may skip Section 6 and Section 7 for the first reading since they do not contribute to the proof of our Main Theorem 1.3 in Section 8. Finally we sketch very briefly in Section 9 how our results extends to the case of polydisks.

The results of this paper, except Section 7 and Section 9, are part of [39] which was submitted in November 2007.

Throughout this paper $\mathbb{E}$ is the standard unit disk, $\mathbb{T}$ is the unit circle and $\mathbb{H}$ is the upper half plane. We denote by $\mathrm{Aut}_{\mathbb{C}}(\mathbb{C})$, $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ and $\mathrm{Aut}_{\mathbb{C}}(\mathbb{P}^1)$ the group of automorphisms of $\mathbb{C}$, $\mathbb{E}$ and $\mathbb{P}^1$ as complex manifolds. The divisor of critical points and the set of critical values of a finite map f are denoted by $\mathfrak{D}_f$ and $\mathfrak{d}_f$. Finally F_r denotes a free group of rank r and $|K|$ is the cardinality of a set K. The set of all homotopy classes of loops with base point q forms the fundamental group of $\mathfrak{M}$ at the point q and is denoted by $\pi_1(\mathfrak{M}, q)$. For path-connected spaces we can write $\pi_1(\mathfrak{M})$ instead of $\pi_1(\mathfrak{M}, q)$ without ambiguity whenever we care about the isomorphism class only. Following usual notations we write E_{ω_1,ω_2} for the elliptic curve $\mathbb{C}/\Lambda_{\omega_1,\omega_2}$ where $\Lambda_{\omega_1,\omega_2}$ is the lattice given by $\mathbb{Z}\omega_1 + \mathbb{Z}\omega_2$.

2 Finite maps and their monodromy groups

Finite maps between Riemann surfaces give a nice category in explaining *fundamental groups detect morphisms* which is one basic principle of Grothendieck's letter. In this section we shall review finite maps by characterizing them in terms of functions fields, fundamental groups and topological monodromy action but we shall start with discussing several examples.

Any annulus $\mathcal{A}$ is conformal to $\mathcal{A}(r,t) = \{z : r < |z| < t \le \infty\}$ with $0 \le r < t \le \infty$. The *modulus* of $\mathcal{A}$, denoted by $\mu(\mathcal{A})$, is defined to be $\ln(t/r)$.

Example 2.1. *If $f : \mathcal{A} \to \mathcal{A}'$ is a finite map between annuli then it is unramified and we have*

$$\mu(\mathcal{A}') = \deg f \cdot \mu(\mathcal{A}).$$

Fatou proved in [16] the following (in an earlier paper [15], he proved the rationality of finite endomorphisms of $\mathbb{E}$ by Schwarz reflection principle).

Example 2.2 (Fatou's theorem)**.** *If f is a holomorphic map from $\mathbb{E}$ to $\mathbb{E}$ then it is finite if and only if it is given by a finite Blaschke product*

$$f(z) = \xi \prod_{i=1}^{n} \frac{z - a_i}{1 - \overline{a_i} z}$$

with $\xi \in \mathbb{T}, n \in \mathbb{N}$ and $a_i \in \mathbb{E}$.

For the details of finite maps between $\mathbb{C}$, $\mathbb{E}$ or annuli we refer the reader to [32, p.211-217]. In contrast there are no finite maps between $\mathbb{C}$ and $\mathbb{E}$. This is a consequence of Liuville's theorem and the following

Lemma 2.3. *If f is a finite map from $\mathbb{E}$ to $\mathfrak{N}$ then $\mathfrak{N}$ is biholomorphic to $\mathbb{E}$.*

Proof. Let $\overline{\mathfrak{N}}$ be the universal covering of $\mathfrak{N}$. By [8, p.99] we deduce that the finiteness of f implies the finiteness of the lifting map $\overline{f} : \mathbb{E} \to \overline{\mathfrak{N}}$ and of the projection map $\pi : \overline{\mathfrak{N}} \to \mathfrak{N}$ which leads to $\pi_1(\mathfrak{N}) = 1$. Firstly $\mathfrak{N}$ cannot be the Riemann's sphere because there is no proper map from a non-compact space to a compact space. Now we claim that $\mathfrak{N}$ cannot be the complex plane and therefore it has to be biholomorphic to the unit disk as claimed. Otherwise we may assume that f is a finite map from $\mathbb{E}$ to $\mathbb{C}$ and then a bounded holomorphic function of $\mathbb{E}$ descends to a bounded holomorphic function of $\mathbb{C}$ by taking the symmetric product. This will imply that there is a non-constant bounded holomorphic function on $\mathbb{C}$, which is impossible. $\square$

From the point of view of birational geometry one can put Ritt's theory into a general geometric setting by employing the analytic function field $\mathbb{C}(\mathfrak{N})$ of a Riemann surface $\mathfrak{N}$. It is known that $\mathfrak{N}$ is uniquely determined by $\mathbb{C}(\mathfrak{N})$ (see for instance [3]) and from [37] that finite maps $f : \mathfrak{M} \to \mathfrak{N}$ are in one-to-one correspondence with finite fields extensions $\mathbb{C}(\mathfrak{N}) \subset \mathcal{K}$ given by $f \mapsto f^\sharp : \mathbb{C}(\mathfrak{N}) \to \mathbb{C}(\mathfrak{M})$. Alternatively we can also characterize finite maps in terms of the fundamental group.

Theorem 2.4 ([37])**.** *Let Σ be a discrete subset of $\mathfrak{N}$ and $q \notin \Sigma$ a point in $\mathfrak{N}$. There is a one-to-one correspondence between finite maps $f : (\mathfrak{M}, p) \to (\mathfrak{N}, q)$ of degree n with $\mathfrak{d}_f \subset \Sigma$ and subgroups H of $\pi_1(\mathfrak{N} \setminus \Sigma, q)$ of index n given by $f \mapsto H = \pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma), p)$.*

We call two proper factorizations

$$\mathfrak{M} \stackrel{\varphi_1}{\to} \mathfrak{T}_1 \stackrel{\varphi_2}{\to} \mathfrak{T}_2 \to \cdots \to \mathfrak{T}_{r-1} \stackrel{\varphi_r}{\to} \mathfrak{N}$$

and

$$\mathfrak{M} \stackrel{\psi_1}{\to} \mathfrak{V}_1 \stackrel{\psi_2}{\to} \mathfrak{V}_2 \to \cdots \to \mathfrak{V}_{s-1} \stackrel{\psi_s}{\to} \mathfrak{N}$$

equivalent if $r = s$ and there exist biholomorphic maps ϕ_i such that the diagram

$$\begin{array}{ccccccccc}
\mathfrak{M} & \stackrel{\varphi_1}{\to} & \mathfrak{T}_1 & \stackrel{\varphi_2}{\to} & \mathfrak{T}_2 & \to \cdots \to & \mathfrak{T}_{r-1} & \stackrel{\varphi_r}{\to} & \mathfrak{N} \\
\downarrow id & & \downarrow \phi_1 & & \downarrow \phi_2 & & \downarrow \phi_{r-1} & & \downarrow id \\
\mathfrak{M} & \stackrel{\psi_1}{\to} & \mathfrak{V}_1 & \stackrel{\psi_2}{\to} & \mathfrak{V}_2 & \to \cdots \to & \mathfrak{V}_{r-1} & \stackrel{\psi_r}{\to} & \mathfrak{N}
\end{array}$$

commutes.

Corollary 2.5. *Let Σ be a discrete subset of $\mathfrak{N}$, p a point in $\mathfrak{M}$, $q \notin \Sigma$ a point in $\mathfrak{N}$ and $f : (\mathfrak{M}, p) \to (\mathfrak{N}, q)$ a finite map with $\mathfrak{d}_f$ contained in Σ. There is a one-to-one correspondence between proper factorizations of f and proper chains of groups between $\pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma), p)$ and $\pi_1(\mathfrak{N} \setminus \Sigma, q)$.*

Let f be a finite map from $\mathfrak{M}$ to $\mathfrak{N}$ of degree n and $q \notin \mathfrak{d}_f$ a point in $\mathfrak{N}$. If we write $f^{-1}(q) = \{p_1, p_2, \cdots, p_n\}$ then for all $\alpha \in \pi_1(\mathfrak{N} \setminus \mathfrak{d}_f, q)$ and for all $i \in \{1, 2, \cdots, n\}$ there is a uniquely determined $(p_i)^\alpha \in f^{-1}(q)$ and a path β from p_i to $(p_i)^\alpha$, unique up to homotopy, such that $f_*\beta = \alpha$. There is a uniquely defined $\rho(\alpha) \in S_n$ such that $(p_i)^\alpha = p_{i\rho(\alpha)}$ for all $1 \leq i \leq n$ and we call the group homomorphism $\rho : \pi_1(\mathfrak{N} \setminus \mathfrak{d}_f, q) \to S_n$ the *monodromy* and the image of ρ the *monodromy group* of f. The monodromy group of f is transitive because $\mathfrak{M}$ is connected. We shall need the following useful remark which complements Theorem 2.4:

$$\pi_1(\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f), p) = \{\alpha \in \pi_1(\mathfrak{N} \setminus \mathfrak{d}_f, q) : p^\alpha = p\}. \tag{1}$$

Here we write p^α instead of $(p)^\alpha$.

If f is the Chebyshev polynomial T_n then

$$\mathfrak{d}_f = \{-1, 1\} \text{ if } n \geq 3, \{-1\} \text{ if } n = 2 \text{ or } \emptyset \text{ if } n = 1.$$

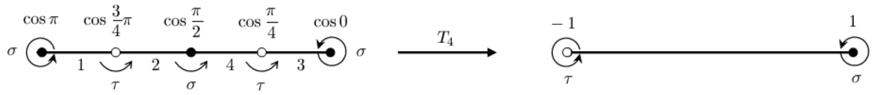
In any case we can look at the monodromy representation of $\pi_1(\mathbb{C} \setminus \{-1, 1\})$ which is a free group of rank 2 generated by σ and τ with σ and τ represented by closed paths around -1 and 1 with counterclockwise orientation. We claim that if $n = 2k$ then

$$\begin{aligned} \rho(\sigma) &= (2,\, 2k)\,(3,\, 2k-1) \cdots (k,\, k+2) \\ \rho(\tau) &= (2,\, 1)\,(3,\, 2k) \cdots (k+1,\, k+2) \end{aligned}$$

and if $n = 2k + 1$ then

$$\begin{aligned} \rho(\sigma) &= (2,\, 2k+1)\,(3,\, 2k) \cdots (k+1,\, k+2) \\ \rho(\tau) &= (2,\, 1)\,(3,\, 2k+1) \cdots (k+1,\, k+3). \end{aligned}$$

We shall call a group homomorphism $\rho : F_2 = \langle \sigma, \tau \rangle \to S_n$ a *Chebyshev representation* if it agrees with the one described as above. For instance, the monodromy of T_4 is illustrated by the following figure.



$1^\sigma = 1, \quad 2^\sigma = 4, \quad 3^\sigma = 3, \quad 1^\tau = 2, \quad 4^\tau = 3$

Figure 1: Chebyshev representation.

Proof of the claim. Since this fact is well-known, we only verify it in the case $n = 4$. It is easily checked that under the polynomial map T_4 the preimage of the closed interval $[-1, 1]$ is $[-1, 1]$, the preimage of the point -1 is $\{\cos(3\pi/4), \cos(\pi/4)\}$ and the preimage of the point 1 is $\{\cos\pi, \cos(\pi/2), \cos 0\}$. We mark the 4 copies of the preimage of open interval $(-1,1)$ with $\{1,2,3,4\}$ as in Figure 1 and then see that 1 goes to 2 under the action of τ, 2 goes to 1, 3 goes to 4 and 4 goes to 3. This gives $\mu(\tau)=(1,2)(3,4)$ and similarly $\mu(\sigma)=(2,4)$. $\square$

It is known that a finite map can be uniquely recovered from its monodromy by virtue of the so-called 'Schere und Kleister' surgery [37, p.41] and this leads to the following restatement of Theorem 2.4.

Theorem 2.6 (Riemann's existence Theorem)**.** *Let $\mathfrak{N}$ be a Riemann surface, Σ a discrete subset in $\mathfrak{N}$ and $\rho : \pi_1(\mathfrak{N}\backslash\Sigma) \to S_n$ a transitive representation. There exists a unique Riemann surface $\mathfrak{M}$ and a finite map f from $\mathfrak{M}$ to $\mathfrak{N}$ with the monodromy of f given by ρ.*

The uniqueness part of the above theorem implies that if finite maps $f : \mathfrak{M} \to \mathfrak{N}$ and $g : \mathfrak{T} \to \mathfrak{N}$ have the same monodromy then there exists a biholomorphic map $\phi : \mathfrak{M} \to \mathfrak{T}$ making the diagram

$$\begin{array}{ccc} \mathfrak{M} & \stackrel{\phi}{\to} & \mathfrak{T} \\ \downarrow_f & & \downarrow_g \\ \mathfrak{N} & \stackrel{id}{\to} & \mathfrak{N} \end{array}$$

commutative.

Remark 2.7. *As a permutation group, the monodromy group of a finite map $f : (\mathfrak{M},p) \to (\mathfrak{N},q)$ is isomorphic to the image of the action of $\pi_1(\mathfrak{N}\backslash\mathfrak{d}_f,q)$ on the coset space $\pi_1(\mathfrak{M}\setminus f^{-1}(\mathfrak{d}_f),p)\backslash\pi_1(\mathfrak{N}\setminus\mathfrak{d}_f,q)$ [37, p.41]. It is also isomorphic to the image of the action of $Gal(K/\mathbb{C}(\mathfrak{N}))$ on the coset space $Gal(K/\mathbb{C}(\mathfrak{M}))\backslash Gal(K/\mathbb{C}(\mathfrak{N}))$ [38, Theorem 5.14], where K is any Galois extension of $\mathbb{C}(\mathfrak{N})$ which contains $\mathbb{C}(\mathfrak{M})$.*

We call a Riemann surface $\mathfrak{M}$ *finite* if $\pi_1(\mathfrak{M})$ is finitely generated. By Ahlfors finiteness theorem [1], this is equivalent to saying that $\mathfrak{M}$ is homeomorphic to a compact Riemann surface with finitely many disks and points deleted. We shall make use of the following version of Riemann-Hurwitz formula.

Lemma 2.8. *Let $\mathfrak{N}$ be a finite Riemann surface. If there exists a finite map f from $\mathfrak{M}$ to another Riemann surface $\mathfrak{N}$ such that $\deg\mathfrak{d}_f<\infty$ then $\mathfrak{M}$ is also finite and*

$$\deg\mathfrak{D}_f = \deg f\cdot\chi_{\mathfrak{N}} - \chi_{\mathfrak{M}}, \tag{2}$$

where $\chi_{\mathfrak{M}}$ and $\chi_{\mathfrak{N}}$ are the Euler characteristic of $\mathfrak{M}$ and $\mathfrak{N}$ respectively.

We shall prove Lemma 2.8 by Schreier's Index Formula applied to fundamental groups and gives an example in explaining "fundamental groups detects morphisms".

Theorem 2.9 (Schreier's Index Formula). *If G is a subgroup of F_r with index i then G is a free group with rank*

$$r_G = i(r-1)+1.$$

Proof of Lemma 2.8. Let $\Sigma \subset \mathfrak{N}$ be a nonempty set with $\mathfrak{d}_f \subset \Sigma$ and $|\Sigma| = n$. We shall calculate $\pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma))$ in two different ways. By elementary topology $\pi_1(\mathfrak{N}\backslash\Sigma) = F_{n+1-\chi_{\mathfrak{N}}}$ and by Theorem 2.4 we have $\deg f = [\pi_1(\mathfrak{N}\backslash\Sigma) : \pi_1(\mathfrak{M}\backslash f^{-1}(\Sigma))]$. Schreier's index formula implies that

$$\pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma)) = F_{\deg f(n-\chi_{\mathfrak{N}})+1}. \tag{3}$$

The map $i_* : \pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma)) \to \pi_1(\mathfrak{M})$ obtained from the inclusion map i is surjective, and so $\mathfrak{M}$ is also finite. Elementary topology again gives

$$\pi_1(\mathfrak{M} \setminus f^{-1}(\Sigma)) = F_{n\deg f - \deg \mathfrak{R}_f + 1 - \chi_{\mathfrak{M}}}. \tag{4}$$

Using the main theorem of finitely generated abelian groups we see that $F_n \sim F_m$ implies that $n = m$. A comparison of (3) and (4) leads to

$$\deg f(n-\chi_{\mathfrak{N}}) + 1 = n \deg f - \deg \mathfrak{D}_f + 1 - \chi_{\mathfrak{M}},$$

and hence $\deg \mathfrak{D}_f = \deg f \cdot \chi_{\mathfrak{N}} - \chi_{\mathfrak{M}}$. □

3 Ritt's first two theorems

In this section we give a proof of Theorem 1.1 and of Theorem 1.2. Even though our proof carries no essentially new ingredients compared with Ritt's original work [35], we present it with the aim to clarify that Ritt's original ideas extend to the more general category. Moreover, a number of consequences which results from the proofs are needed to prove our Main Theorem 1.3. Notice that a topological version of Theorem 1.1 was already discussed in [24, p.65].

Proof of Theorem 1.1. Choose $q \notin \mathfrak{d}_f$ in $\mathfrak{N}$ and $p \in \mathfrak{M}$ with $f(p) = q$ then we deduce from Corollary 2.5 that f is prime if and only if $\pi_1(\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f), p)$ is a maximal subgroup of $\pi_1(\mathfrak{N} \setminus \mathfrak{d}_f, q)$ and this is equivalent to $\pi_1(\mathfrak{N} \setminus \mathfrak{d}_f)$ acting primitively on $\pi_1(\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f))\backslash\pi_1(\mathfrak{N} \setminus \mathfrak{d}_f)$ (see for instance [10]) and now Remark 2.7 gives the desired result. □

We shall recall some basic lattice theory and we shall follow the notations $x < y$, $x \prec y$, $x \vee y$ and $x \wedge y$ as described in [7]. A lattice $\mathfrak{L}$ is said to satisfy the *Jordan-Dedekind chain condition* if the length of maximal proper chains depends only on the endpoints. We say that $\mathfrak{L}$ is of *locally finite* if every interval of $\mathfrak{L}$ is of finite length. We call $\mathfrak{L}$ *modular* if

$$x \leq z \Rightarrow x \vee (y \wedge z) = (x \vee y) \wedge z \quad \forall\, y \in \mathfrak{L}.$$

The following modular lattices play an important role in Ritt's theory.

Example 3.1. *Let $\mathfrak{L}_n = \{\, t \in \mathbb{N} : t \,|\, n \,\}$ so that $i \leq j$ if and only if $i \,|\, j$. Then $(\, \mathfrak{L}_n \,;\, \leq \,)$ is a lattice and $x \vee y = lcm(x, y), x \wedge y = \gcd(x, y)$. This lattice of devisors of n is modular and any sublattice $\mathfrak{F}$ of $(\, \mathfrak{L}_n \,;\, \leq \,)$ is also modular.*

If a locally finite lattice $\mathfrak{L}$ is modular, then it satisfies the Jordan-Dedekind chain condition. Furthermore there is a *dimension function* $d : \mathfrak{L} \to \mathbb{Z}$ such that $x \prec y$ if and only if $x < y$ and $d\,(y) = d\,(x) + 1$ for all $x, y \in \mathfrak{L}$. In addition we have $d\,(x) + d\,(y) = d\,(x \vee y) + d\,(x \wedge y)$. Ritt [35] proved an important property for sublattices of $(\, \mathfrak{L}_n \,;\, \leq \,)$. The following proposition extends Ritt's result to general modular lattices.

Proposition 3.2. *Let $\mathfrak{L}$ be a locally finite modular lattice, $a, b \in \mathfrak{L}$ with $a < b$ and $\mathcal{C}, \mathcal{C}'$ maximal proper chains of $\mathfrak{L}$ with the same endpoints a and b. There exists $m \geq 0$ and a sequence of maximal proper chains $\mathcal{C}_i$, $0 \leq i \leq m$, with endpoints a and b such that $\mathcal{C}_0 = \mathcal{C}$, $\mathcal{C}_m = \mathcal{C}'$ and $\mathcal{C}_i$ and $\mathcal{C}_{i+1}$ differ in only one element.*

Proof. We write $\mathcal{C}$ and $\mathcal{C}'$ as

$$\begin{aligned} \mathcal{C} : a = \quad & x_0 \prec x_1 \prec x_2 \prec x_3 \prec \cdots \prec x_n \quad &= b, \\ \mathcal{C}' : a = \quad & y_0 \prec y_1 \prec y_2 \prec y_3 \prec \cdots \prec y_n \quad &= b, \end{aligned}$$

choose a dimension function d and prove the claim by induction. If $n = 2$ nothing requires a proof. Assume the claim holds for all $2 \leq n \leq k - 1$, we will prove it for $n = k$.

If $x_1 = y_1$ we apply the induction assumption to $\mathcal{B} : x_1 \prec x_2 \prec x_3 \prec \cdots \prec x_k = b$ and $\mathcal{B}' : y_1 \prec y_2 \prec y_3 \prec \cdots \prec y_k = b$ and this proves the proposition in the case $n < k$ or $n = k, x_1 = y_1$.

It remains the case that $n = k, y_1 \neq x_1$. We first show that $y_1 \nleq x_1$. If not then $y_1 < x_1$ and this leads to $d(a) < d(y_1) < d(x_1)$, a contradiction to $d(x_1) = d(a) + 1$. Since $y_1 \leq x_k = b$, there exists $1 \leq i \leq k - 1$ such that $y_1 \nleq x_i,\ y_1 \leq x_{i+1}$. If $i = 1$ we put $\mathcal{C}_0 = \mathcal{C}$ and define $\mathcal{C}_1 : x_0 \prec y_1 \prec x_2 \prec x_3 \prec \cdots \prec x_k$. To go from $\mathcal{C}_1$ to $\mathcal{C}'$ we note that here the case where $n = k$ and $x_1 = y_1$ applies and we are done.

Now we assume the proposition holds when $1 \leq i \leq l-1$ and we prove it for $i = l$. Since

$$\begin{aligned} d(y_1 \vee x_{l-1}) &= d(y_1) + d(x_{l-1}) - d(y_1 \wedge x_{l-1}) \\ &= d(y_1) + d(x_{l-1}) - d(x_0) \\ &= d(x_{l-1}) + 1 \end{aligned}$$

and since $y_1 \leq x_{l+1}$ implies $x_{l-1} \leq y_1 \vee x_{l-1} \leq x_{l+1}$, we conclude that $x_{l-1} \prec y_1 \vee x_{l-1} \prec x_{l+1}$. This shows that we can choose $\mathcal{C}_0 = \mathcal{C}$ and $\mathcal{C}_1 = x_0 \prec x_1 \prec x_2 \cdots \prec x_{l-1} \prec y_1 \vee x_{l-1} \prec x_{l+1} \prec \cdots \prec x_k$. To go from $\mathcal{C}_1$ to $\mathcal{C}'$ we see that the case $i = l-1$ applies and we are done. $\square$

As an illustration we give the following

Example 3.3. *Let $\mathfrak{F}$ be a sublattice of $\mathfrak{L}_n$ and $\mathcal{C}, \mathcal{C}'$ maximal proper chains of $\mathfrak{F}$ with endpoints* 1 *and n. There exists $m \in \mathbb{N}$ and a sequence of maximal proper chains $\mathcal{C}_i$, $0 \leq i \leq m$, with endpoints* 1 *and n such that $\mathcal{C}_0 = \mathcal{C}$, $\mathcal{C}_m = \mathcal{C}'$ and any two consecutive ones $\mathcal{C}_j$ and $\mathcal{C}_{j+1}$ differ only in one element. This means that we can write $\mathcal{C}_j$ as $\cdots \prec a_i \prec a_{i+1} \prec a_{i+2} \prec \cdots$ and $\mathcal{C}_{j+1}$ as $\cdots \prec a_i \prec a'_{i+1} \prec a_{i+2} \prec \cdots$ respectively. Both two chains are proper and $a_{i+1} \neq a'_{i+1}$, as a consequence we have*

$$\gcd\left(\frac{a_{i+1}}{a_i}, \frac{a'_{i+1}}{a_i}\right) = 1, \quad \gcd\left(\frac{a_{i+2}}{a_{i+1}}, \frac{a_{i+2}}{a'_{i+1}}\right) = 1$$

or equivalently

$$\frac{a_{i+1}}{a_i} = \frac{a_{i+2}}{a'_{i+1}}, \ \frac{a_{i+2}}{a_{i+1}} = \frac{a'_{i+1}}{a_i}, \ \gcd\left(\frac{a_{i+1}}{a_i}, \frac{a_{i+2}}{a_{i+1}}\right) = 1. \tag{5}$$

For the proof of Theorem 1.2 we also need

Dedekind's Modular Law ([36]) *Let G be a group and $H \leq K \leq G$ and $L \leq G$ subgroups. Then we have $(LH) \cap K = (L \cap K)H$.*

Proof of Theorem 1.2. We write $n = \deg f$, $\alpha(0) = q$ and choose $f^{-1}(q) = \{p = p_1, p_2, \cdots, p_n\}$. According to Corollary 2.5 it suffices to prove that the lattice $\mathfrak{L}$ consisting of all intermediate groups between $G = \pi_1(\mathfrak{N} \setminus \mathfrak{d}_f, q)$ and $H = \pi_1(\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f), p)$ is modular. For $K \in \mathfrak{L}$ we write $K_\alpha = K \cap \langle\alpha\rangle$. Trivially $G_\alpha = \langle\alpha\rangle$. Moreover by the transitivity of the action of α on $f^{-1}(q)$ we have

$$H_\alpha = H \cap \langle\alpha\rangle \stackrel{(1)}{=} \{\beta \in \langle\alpha\rangle : p_1 = p = p^\beta = p_1^\beta = p_{1^\beta}\} = \{\beta \in \langle\alpha\rangle : 1^\beta = 1\} = \langle\alpha^n\rangle$$

The fact that $H \cap \langle\alpha\rangle = \langle\alpha^n\rangle$ immediately leads to $\alpha^i H \neq \alpha^j H$ for all $0 \leq i < j \leq n-1$. This together with $[G : H] = n$ leads to $G = \cup_{i=0}^{n-1} \alpha^i H$ and in particular

$G = \langle\alpha\rangle H$. Since $\mathfrak{L}_n$ is isomorphic to the lattice consisting of intermediate groups between $\langle\alpha\rangle$ and $\langle\alpha^n\rangle$, we shall treat them equally and our above discussion defines a map $g : \mathfrak{L} \to \mathfrak{L}_n$ by $g(K) = K_\alpha$. By Dedekind's Modular Law and by $G = \langle\alpha\rangle H$ we have

$$K_\alpha H = (\langle\alpha\rangle \cap K)H = \langle\alpha\rangle H \cap K = G \cap K = K. \tag{6}$$

This implies immediately that g is injective. Since K is a group we deduce from (6) that $K_\alpha H$ is also a group and this leads to

$$K_\alpha H = HK_\alpha. \tag{7}$$

To prove that g is a lattice morphism it suffice to verify that $K_\alpha \cap M_\alpha = (K \cap M)_\alpha$ and $\langle K, M\rangle_\alpha = \langle K_\alpha, M_\alpha\rangle$ for all $K, M \in \mathfrak{L}$. The former is trivial since

$$K_\alpha \cap M_\alpha = K \cap \langle\alpha\rangle \cap M \cap \langle\alpha\rangle = (K \cap M) \cap \langle\alpha\rangle = (K \cap M)_\alpha.$$

By $\langle K, M\rangle = \langle K_\alpha H, M_\alpha H\rangle \overset{(7)}{=} HK_\alpha M_\alpha$ and by $K_\alpha M_\alpha = M_\alpha K_\alpha$ the latter follows from

$$\langle K, M\rangle_\alpha = \langle K, M\rangle \cap \langle\alpha\rangle = HK_\alpha M_\alpha \cap \langle\alpha\rangle = (H \cap \langle\alpha\rangle)K_\alpha M_\alpha = K_\alpha M_\alpha,$$

where the second last equality relies on Dedekind's modular law. We have proved that g is an injective lattice morphism and this gives that $\mathfrak{L} \simeq g(\mathfrak{L})$ and the latter is a sublattice of $\mathfrak{L}_n$. We conclude from Example 3.1 that $\mathfrak{L}$ is modular. $\square$

A much shorter proof exists, but our proof gives more information. In particular it implies that $\mathfrak{L}$ is a sublattice of $\mathfrak{L}_n$. According to (5), we can pass from one maximal factorization of f to another with each step given by a solution $(\phi_i, \phi_{i+1}, \phi'_i, \phi'_{i+1})$ to the two finite maps equation

$$\phi_i \circ \phi_{i+1} = \phi'_i \circ \phi'_{i+1}, \deg \phi_i = \deg \phi'_{i+1},\ \gcd(\deg \phi_i, \deg \phi_{i+1}) = 1. \tag{8}$$

This functional equation in polynomials is a major difficulty solved in [35] and we shall solve this functional equation in finite Blaschke products.

4 Deformation

We shall make use of the fact that finite endomorphisms of $\mathbb{E}$ can be deformed to finite endomorphisms of $\mathbb{C}$. This follows from Riemann's covering principle, for which we refer to [2, p.119-120]. A Riemann surface is a pair $(\mathfrak{W}, \Phi)$ with $\mathfrak{W}$ a connected Hausdorff space and Φ a complex structure, see [2, p.144]. However we shall simply write $\mathbb{E}$ and $\mathbb{C}$ when Φ is canonical.

Theorem 4.1 (Riemann's covering principle)**.** *If $f : \mathfrak{W}_1 \to \mathfrak{W}_2$ is a covering surface and if $\mathfrak{W}_2$ admits a complex structure Φ_2 then there exists a unique complex structure Φ_1 on $\mathfrak{W}_1$ which makes f a holomorphic map from $(\mathfrak{W}_1, \Phi_1)$ to $(\mathfrak{W}_2, \Phi_2)$.*

Let f be a finite map from $\mathbb{E}$ to $\mathbb{E}$ and i_0 a homeomorphism from $\mathbb{E}$ to $\mathbb{C}$. The canonical complex structure on $\mathbb{C}$ induces a new complex structure Φ_0 on $\mathbb{E}$ and we obtain a new Riemann surface $(\mathbb{E}, \Phi_0)$. By Theorem 4.1 applied to $f : \mathbb{E} \to (\mathbb{E}, \Phi_0)$ there exists a Riemann surface $(\mathbb{E}, \Phi_1)$ such that f is a holomorphic map from $(\mathbb{E}, \Phi_1)$ to $(\mathbb{E}, \Phi_0)$. Consequently there exists a holomorphic map $(i_1, i_0)_* f : (\mathbb{E}, \Phi_1) \to \mathbb{C}$ such that $(i_1, i_0)_* f \circ i_1 = i_0 \circ f$ where i_1 is the topological identity map $i_1 : \mathbb{E} \to (\mathbb{E}, \Phi_1)$. We shall call i_1 a *lifting* of i_0 by f and $(i_1, i_0)_* f$ a *descent* of f by the pair (i_1, i_0).

$$\begin{array}{ccc} \mathbb{E} & \stackrel{f}{\longrightarrow} & \mathbb{E} \\ \downarrow i_1 & & \downarrow i_0 \\ (\mathbb{E}, \Phi_1) & \stackrel{(i_1,i_0)_* f}{\longrightarrow} & \mathbb{C} \end{array}$$

The uniqueness part of Theorem 4.1 shows that if i_1, i_1' are two liftings of i_0 then there exists a holomorphic isomorphism σ between $(\mathbb{E}, \Phi_1)$ and $(\mathbb{E}, \Phi_1')$ such that $\sigma \circ i_1 = i_1'$. The classical uniformization theorem for simply connected Riemann surfaces together with Lemma 2.3 shows that $(\mathbb{E}, \Phi_1)$ and $(\mathbb{E}, \Phi_1')$ must be biholomorphic to $\mathbb{C}$. To sum up we may state the following

Corollary 4.2. *Let f be a finite map from $\mathbb{E}$ to $\mathbb{E}$, i_0 a homeomorphism from $\mathbb{E}$ to $\mathbb{C}$ and $i_1, i_1' : \mathbb{E} \to \mathbb{C}$ liftings of i_0 by f. There exists σ in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{C})$ such that $i_1' = \sigma \circ i_1$ and $(i_1', i_0)_* f \circ \sigma = (i_1, i_0)_* f$.*

This can be illustrated by the following diagram

$$\begin{array}{ccccc} & & \mathbb{E} & \stackrel{f}{\longrightarrow} & \mathbb{E} \\ & \swarrow i_1' & \downarrow i_1 & & \downarrow i_0 \\ \mathbb{C} & \stackrel{\sigma = az+b}{\longleftarrow} & \mathbb{C} & \stackrel{(i_1,i_0)_* f}{\longrightarrow} & \mathbb{C} \end{array}$$

where $(i_1, i_0)_* f$ is a finite endomorphism of $\mathbb{C}$ and therefore is a polynomial. Our construction of liftings is functorial.

Proposition 4.3. *Let f_1, f_2 be finite endomorphisms of $\mathbb{E}$, i_0 a homeomorphism from $\mathbb{E}$ to $\mathbb{C}$ and $f = f_1 \circ f_2$. If i_1 and i_2 are lifting of i_0 by f_1 and of i_1 by f_2 respectively then i_2 is a lifting of i_0 by f and $(i_2, i_0)_* f = (i_1, i_0)_* f_1 \circ (i_2, i_1)_* f_2$ is a composition of polynomials.*

$$\begin{array}{ccccc} \mathbb{E} & \xrightarrow{f_2} & \mathbb{E} & \xrightarrow{f_1} & \mathbb{E} \\ \downarrow i_2 & & \downarrow i_1 & & \downarrow i_0 \\ \mathbb{C} & \xrightarrow{(i_2,i_1)_* f_2} & \mathbb{C} & \xrightarrow{(i_1,i_0)_* f_1} & \mathbb{C} \end{array}$$

Proof. By definition the maps $i_0 \circ f_1 \circ i_1^{-1}$ and $i_1 \circ f_2 \circ i_2^{-1}$ are both holomorphic and by Theorem 4.1 it suffices to show $i_0 \circ f \circ i_2^{-1}$ is holomorphic. This follows from $i_0 \circ f \circ i_2^{-1} = (i_0 \circ f_1 \circ i_1^{-1}) \circ (i_1 \circ f_2 \circ i_2^{-1})$. □

In general, compared with finite Blaschke products, polynomials are easier to deal with since much more algebraic techniques (such as the place at infinity) are available.

5 Chebyshev-Blaschke products

In this section we shall construct *Chebyshev-Blaschke products* using the geometric monodromy action. If a, b is a pair of distinct elements in $\mathbb{E}$ then the group $\pi_1(\mathbb{E} \setminus \{a, b\})$ can be generated by two elements σ and τ with σ and τ represented by closed paths around a and b with counterclockwise orientation.

Lemma 5.1. *For $n \in \mathbb{N}$ there exists a finite endomorphism $f_{n,a,b}$ of $\mathbb{E}$ such that*
1) $\deg f_{n,a,b} = n$, $\mathfrak{d}_{f_{n,a,b}} = \{a, b\}$ (if $n > 2$) or $\{a\}$ (if $n = 2$) or $\emptyset$ (if $n = 1$).
2) The monodromy representation $\rho : \langle \sigma, \tau \rangle \to S_n$ is a Chebyshev representation.
The map $f_{n,a,b}$ is unique up to composition on the right with an element in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$.

Proof. Theorem 2.6 gives a finite map $f : \mathfrak{M} \to \mathbb{E}$ which satisfies the monodromy condition. By Lemma 2.8 a direct calculation leads to $\chi(\mathfrak{M}) = 1$ and it follows from elementary topology that $\mathfrak{M}$ is either $\mathbb{C}$ or $\mathbb{E}$. Liouville's Theorem rules out the possibility of $\mathbb{C}$ and the uniqueness part of Theorem 2.6 completes the proof. □

We will call those $f_{n,a,b}$ *Chebyshev-Blaschke products*. In order to describe normalized forms of $f_{n,a,b}$, we denote by $\gamma(t)$ for any positive real number t the unique number in $(0, 1)$ such that $\mu(\mathbb{E} \setminus [-\gamma(t), \gamma(t)]) = t$. Given $t > 0$ and $n \in \mathbb{N}$ we take $a = -\gamma(nt)$ and $b = \gamma(nt)$.

Proposition 5.2. *For all positive real number t and positive integer n there is a finite endomorphism f of $\mathbb{E}$ with $\deg f = n$ which satisfies*

1) $\mathfrak{d}_f = \{a, b\}$ (if $n > 2$) or $\{a\}$ (if $n = 2$) or $\emptyset$ (if $n = 1$).

2) $f^{-1}([-\gamma(nt), \gamma(nt)]) = [-\gamma(t), \gamma(t)]$ and $f(\gamma(t)) = \gamma(nt)$.

3) The monodromy representation $\rho : \langle \sigma, \tau \rangle \to S_n$ is a Chebyshev representation.

Before the proof we recall some geometry and topology. The isometry group $\mathrm{Isom}(\mathbb{E}, ds)$ of $\mathbb{E}$ with respect to the Poincaré metric ds is given by the semidirect product $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E}) \rtimes \langle i \rangle$, where i is complex conjugation. We write $\mathrm{Isom}^+(\mathbb{E}, ds)$ for the set of holomorphic automorphisms and $\mathrm{Isom}^-(\mathbb{E}, ds)$ for the antiholomorphic ones. The fixed point set $\mathrm{Fix}(\iota)$ of an element ι in $\mathrm{Isom}(\mathbb{E}, ds)$ is either empty, a point, a geodesic line or $\mathbb{E}$. Let f be a finite map from $\mathfrak{M}$ to $\mathfrak{N}$, t a homeomorphism from $\mathfrak{N}$ to $\mathfrak{N}$, $q \notin \mathfrak{d}_f$ a point in $\mathfrak{N}$ and p_1, p_2 points in $\mathfrak{M}$ with $f(p_i) = q$. Elementary topology shows that the map t lifts to a homeomorphism $\iota : (\mathfrak{M}, p_1) \to (\mathfrak{M}, p_2)$ making the following diagram

$$\begin{array}{ccc} \mathfrak{M} & \xrightarrow{\iota} & \mathfrak{M} \\ \downarrow f & & \downarrow f \\ \mathfrak{N} & \xrightarrow{t} & \mathfrak{N} \end{array}$$

commutative if and only if t restricts to a bijection on $\mathfrak{d}_f$ and if $(t \circ f)_*(\,\pi_1(\,\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f), p_1\,)\,) = f_*(\,\pi_1(\,\mathfrak{M} \setminus f^{-1}(\mathfrak{d}_f), p_2\,)\,)$.

Proof of Proposition 5.2. Lemma 5.1 gives a finite map $f : \mathbb{E} \to \mathbb{E}$ which satisfies 1) and 3). Moreover if we can prove that $f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,)$ is a geodesic segment then 2) is immediately fulfilled by composing f with an element in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$. We only verify this fact for $n = 2k$, since similar considerations apply to $n = 2k + 1$.

By condition f is an unramified map from $\mathbb{E} \setminus f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,)$ to an annulus $\mathbb{E} \setminus [-\gamma(nt), \gamma(nt)]$. This implies that $\mathbb{E} \setminus f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,)$ is an annulus and therefore $f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,)$ is connected.

Choose $q \in (\,-\gamma(nt), \gamma(nt)\,)$ and write $f^{-1}(q) = \{p_1, p_2, \cdots, p_{2k}\}$ with the numbering i chosen such that $p_i^\alpha = p_{i\rho(\alpha)}$ for $1 \le i \le 2k$, $\alpha \in \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,)$. We show now that there is a commutative diagram

$$\begin{array}{ccc} (\mathbb{E}, p_1) & \xrightarrow{\iota} & (\mathbb{E}, p_1) \\ \downarrow f & & \downarrow f \\ (\mathbb{E}, q) & \xrightarrow{i} & (\mathbb{E}, q) \end{array}$$

with an isometry ι in $I(\mathbb{E})$ such that $f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,) \subset \mathrm{Fix}(\iota)$. As a consequence $f^{-1}(\,[-\gamma(nt), \gamma(nt)]\,)$ will be a geodesic segment. By the remark before it suffices to show that i restricts to a bijection on $\{a, b\}$ and that $(i \circ f)_*(\,\pi_1(\,\mathbb{E} \setminus f^{-1}\{a, b\}, p_1\,)\,) = f_*(\,\pi_1(\,\mathbb{E} \setminus f^{-1}\{a, b\}, p_1\,)\,)$. The involution $i : \mathbb{E} \setminus \{a, b\} \to \mathbb{E} \setminus \{a, b\}$ induces a map $i_* : \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,) \to \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,)$. The base point q of σ and τ is on the interval $(-\gamma(nt), \gamma(nt))$ and therefore our involution i on σ and τ simply changes the orientation, and this means that

$$i_*(\sigma) = \sigma^{-1}, \quad i_*(\tau) = \tau^{-1}.$$

By condition that ρ is a Chebyshev representation we have both $\rho(\sigma)$ and $\rho(\tau)$ are of order two and therefore $\rho(i_*(\sigma)) = \rho(\sigma)$ as well as $\rho(i_*(\tau)) = \rho(\tau)$. This gives $\rho \circ i_* = \rho$ on $\langle \sigma, \tau \rangle = \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,)$ which displays as

$$\rho(i_*(\alpha)) = \rho(\alpha), \qquad \forall\ \alpha \in \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,). \tag{9}$$

We use (1) to deduce that

$$i_*(f_*(\pi_1(\mathbb{E} \setminus f^{-1}\{a, b\},\, p_1))) = i_*(\{\alpha : \alpha \in \pi_1(\mathbb{E} \setminus \{a, b\},\, q), 1^{\rho(\alpha)} = 1\})$$

and observe that β to be in the group on the right is equivalent to $i_*^{-1}(\beta)$ to be in the group $\{\alpha : \alpha \in \pi_1(\mathbb{E} \setminus \{a, b\},\, q), 1^{\rho(\alpha)} = 1\}$. Therefore the right hand side equals $\{\alpha : \alpha \in \pi_1(\mathbb{E} \setminus \{a, b\},\, q), 1^{\rho(i_*^{-1}(\alpha))} = 1\}$. Using (9) we find that this is the same as

$$\{\alpha : \alpha \in \pi_1(\mathbb{E} \setminus \{a, b\},\, q), 1^{\rho(\alpha)} = 1\} = f_*(\pi_1(\mathbb{E} \setminus f^{-1}(\{a, b\}),\, p_1)).$$

This shows that the involution i lifts to a homeomorphism $\iota : (\mathbb{E}, p_1) \to (\mathbb{E}, p_1)$ and from the diagram we deduce with elementary topology that

$$\iota(p_1^{\alpha}) = p_1^{i_*(\alpha)}, \qquad \forall\ \alpha \in \pi_1(\,\mathbb{E} \setminus \{a, b\},\, q\,) = \langle \sigma, \tau \rangle\,.$$

In particular $\iota(p_1^{\tau}) = p_1^{i_*(\tau)}$ and therefore $\iota(p_2) = p_2$. Similar arguments show that $\iota(p_j) = p_j$ for all $1 \le j \le 2k$ and we get $f^{-1}(q) \subset \mathrm{Fix}(\iota)$. We differentiate the equation $f(\iota(z)) = \overline{f(z)}$ which follows from the diagram. This implies that $\partial\iota/\partial z = 0$ which means that ι is an antiholomorphic homeomorphism of the unit disk. As a consequence $\iota \in \mathrm{Isom}^-(\mathbb{E}, ds)$ is an isometry and therefore it suffices to prove that $f^{-1}([-\gamma(nt), \gamma(nt)]) \subset \mathrm{Fix}(\iota)$. The paths σ and τ, the preimage p_i and the lift ι vary continuously if q varies continuously in $(-\gamma(nt), \gamma(nt))$. In addition for given f and i the equation $i \circ f = f \circ \iota$ has only finitely many solutions ι in $\mathrm{Isom}^-(\mathbb{E}, ds)$. Indeed choose a fixed point $x \in \mathbb{E}$ then any solution ι takes values at x in a finite set $f^{-1}(i(f(x)))$ and since ι is an antiholomorphic automorphism it is uniquely determined by the image at two distinct points. This shows that there are only finitely many possibilities. We conclude that ι is locally constant and therefore independent of q. This shows that $f^{-1}([-\gamma(nt), \gamma(nt)]) \subset \mathrm{Fix}(\iota)$ as claimed. $\square$

Proposition 5.3. *For all positive real number t and positive integer n there exists a unique finite endomorphism $f_{n,t}$ of the unit disk $\mathbb{E}$ with the property that $f^{-1}([-\gamma(nt), \gamma(nt)]) = [-\gamma(t), \gamma(t)]$ and $f(\gamma(t)) = \gamma(nt)$.*

Proof. The existence of $f_{n,t}$ comes from Proposition 5.2 and therefore it suffices to prove that any two such maps f_1 and f_2 coincide. As a first step we show that $\mathfrak{d}_f \subset \{-\gamma(nt), \gamma(nt)\}$ for f in $\{f_1, f_2\}$.

The map f restricts to finite maps from the annulus $\mathbb{E} \setminus f^{-1}([-\gamma(nt), \gamma(nt)])$ which is $\mathbb{E} \setminus [-\gamma(t), \gamma(t)]$ to the annulus $\mathbb{E} \setminus [-\gamma(nt), \gamma(nt)]$. In Example 2.1 we have seen that such a map is unramified which shows that $\mathbb{E}_f \subset [-\gamma(nt), \gamma(nt)]$. Moreover the moduli of these annuli differ by a factor n and this shows that $\deg f = n$.

Taking $q \in (-\gamma(nt), \gamma(nt))$ and p a point in $f^{-1}(q) \subset (-\gamma(t), \gamma(t))$ we deduce that the preimage of an open neighborhood of q in $(-\gamma(nt), \gamma(nt))$ is an open neighborhood of p in $(-\gamma(t), \gamma(t))$. Consequencely the preimage of two trajectories in $(-\gamma(nt), \gamma(nt))$ at q consists of two trajectories in $(-\gamma(t), \gamma(t))$ at p and this implies that f is unramified at p. This gives that f is unramified over any point q in $(-\gamma(nt), \gamma(nt))$ showing that $\mathfrak{d}_f \subset \{-\gamma(nt), \gamma(nt)\}$ as stated.

To continue with the proof we distinguish between two cases.

Case $n = 2k$.

Because f is an unramified cover of $(-\gamma(nt), \gamma(nt))$ the preimage of $(-\gamma(nt), \gamma(nt))$ under f is a disjoint union of n real 1-dimensional connected curves in $[-\gamma(t), \gamma(t)]$. As such they have to be open intervals of the form (a_i, b_i) or (b_i, a_{i+1}) for $i = 1, \cdots, k$ with $a_1 < b_1 < a_2 < b_2 < \cdots < b_k < a_{k+1}$, $f(a_i) = r(nt)$, $f(b_i) = -r(nt)$ and $\{a_2, \ldots, a_k\} \cup \{b_1, \ldots, b_k\}$ the critical points. This leads to a picture similar to Figure 1, and an argument similar to that given in the proof of the statement there shows that the monodromy representation $\rho : \langle \sigma, \tau \rangle \to S_n$ is a Chebyshev representation. The uniqueness part of Theorem 2.6 leads to the existence of $\iota \in \mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ with $f_1 = f_2 \circ \iota$. Taking inverse images and using that $f_i^{-1}([-\gamma(nt), \gamma(nt)]) = [-\gamma(t), \gamma(t)]$ leads to $\iota([-\gamma(t), \gamma(t)]) = [-\gamma(t), \gamma(t)]$ whence $\iota(\pm\gamma(t)) = \pm\gamma(t)$ or $\iota(\pm\gamma(t)) = \mp\gamma(t)$. In the former case $\iota = id$ and therefore $f_1 = f_2 \circ id = f_2$. In latter case $\iota = -id$, therefore $f_1 = f_2 \circ (-id)$ and finally to conclude $f_1 = f_2$ it suffices to prove $f_2(z) = f_2(-z)$.

Choose $q \in (\,-\gamma(nt), \gamma(nt)\,)$ and write $f_2^{-1}(q) = \{p_1, p_2, \cdots, p_{2k}\}$ with the numbering i chosen such that $p_i^{\alpha} = p_{i\rho(\alpha)}$ for all $1 \leq i \leq 2k$ and for all $\alpha \in \langle \sigma, \tau \rangle$. Similar to the proof of Proposition 5.2, the map $id : (\mathbb{E}, q) \to (\mathbb{E}, q)$ lifts to a map $\iota : (\mathbb{E}, p_1) \to (\mathbb{E}, p_{k+1})$ different from the identity in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ such that $f_2 \circ \iota = id \circ f_2$ and again $\iota([-\gamma(t), \gamma(t)]) = [-\gamma(t), \gamma(t)]$. This together with the property that $\iota \neq id$ implies that $\iota(z) = -z$ and therefore $f_2(z) = f_2(-z)$ as desired.

Case $n = 2k + 1$.

The preimage of $(-\gamma(nt), \gamma(nt))$ is a disjoint union of $n = 2k + 1$ open intervals of the form (a_i, b_i) or (b_j, a_{j+1}) for $i = 1, \cdots, k+1$ or $j = 1, \cdots, k$ with $f(a_i) = \gamma(nt)$, $f(b_i) = -\gamma(nt)$ and $\mathfrak{D}_f = \sum_{i=2}^{k+1}(a_i) + \sum_{i=1}^{k}(b_i)$. Similar considerations to that as above proceed up to there exists ι in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ such that $f_1 = f_2 \circ \iota$ and $\iota(\pm\gamma(t)) =$

$\pm\gamma(t)$. The latter identity implies $\iota = id$ and therefore $f_1 = f_2$ as desired. $\square$

If $n \geq 3$ and if $f_{n,\,a,\,b}$ is the Chebyshev-Blaschke product constructed in Lemma 5.1 then there exist uniquely an element ϵ in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ and a positive real number t such that $\epsilon(a) = -r(nt)$ and $\epsilon(b) = r(nt)$ and now $\epsilon \circ f_{n,\,a,\,b}$ has the same monodromy as the function $f_{n,\,t}$ constructed in Proposition 5.3. Therefore there exists $\varepsilon \in \mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ such that $\epsilon \circ f_{n,\,a,\,b} \circ \varepsilon = f_{n,\,t}$. The maps $f_{n,\,t}$ obtained in this way is called *normalized Chebyshev-Blaschke products.* We sum up with the following corollary

Corollary 5.4. *If f is a finite map from $\mathbb{E}$ to $\mathbb{E}$ with degree at least three and if its monodromy representation is Chebyshev representation then there exist a positive number t and ϵ, ε in* $\mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ *such that*

$$f_{n,\,a,\,b} = \epsilon \circ f_{n,\,t} \circ \varepsilon$$

and this factorization is unique.

Chebyshev Blaschke products have the following special nesting property.

Theorem 5.5. *For all positive real number t and positive integers m and n we have*

$$f_{mn,\,t} = f_{m,\,nt} \circ f_{n,\,t}.$$

Proof. Direct calculation leads to

$$\begin{aligned}
(f_{m,\,nt} \circ f_{n,\,t})^{-1}\,(\,[-\gamma(mnt), \gamma(mnt)]\,) &= f_{n,\,t}^{-1}\,(\,f_{m,\,nt}^{-1}(\,[-\gamma(mnt), \gamma(mnt)]\,)\,) \\
&= f_{n,\,t}^{-1}\,(\,[-\gamma(nt), \gamma(nt)]\,) \\
&= [-\gamma(t), \gamma(t)]
\end{aligned}$$

and from Proposition 5.3 we deduce that $f_{mn,\,t} = f_{m,\,nt} \circ f_{n,\,t}$. $\square$

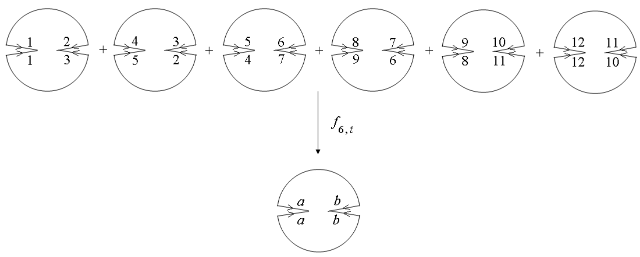
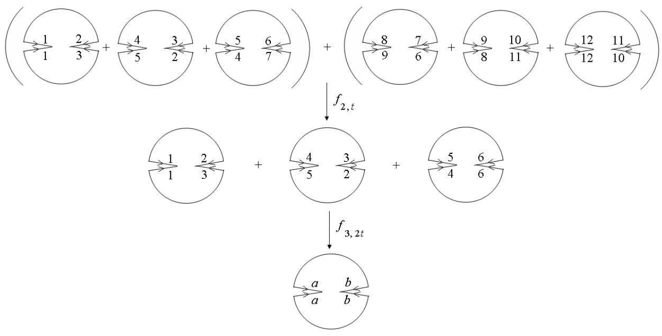


Figure 2: The topology of $f_{6,t}$.

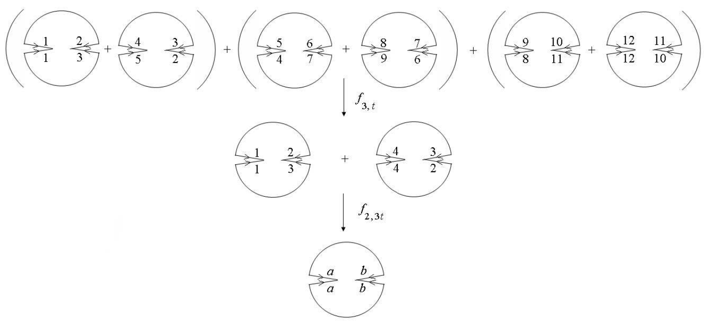
The topological nature of $f_{n,\,t}$ may be illustrated by Riemann's 'Schere und Kleister' surgery applied to copies of the unit disk. If we take $f_{6,\,t}$ as an example then we shall obtain Figure 2.

Figure 3 illustrates the factorization $f_{6,\,t} = f_{3,\,2t} \circ f_{2,\,t}$ and Figure 4 illustrates the factorization $f_{6,\,t} = f_{2,\,3t} \circ f_{3,\,t}$.



$f_{2,t} : 7 \to 6, 8 \to 5, 9 \to 4, 10 \to 3, 11 \to 2, 12 \to 1.$

Figure 3: The first factorization of $f_{6,t}$.



$f_{3,t} : 5 \to 4, 6 \to 3, 7 \to 2, 8 \to 1, 9 \to 1, 10 \to 2, 11 \to 3, 12 \to 4$

Figure 4: The second factorization of $f_{6,t}$.

6 Jacobian elliptic functions

The reader who is only interested in Ritt's theory on the unit disk may read Section 8 first and return to this section and Section 7 later if he wants to know more about Chebyshev-Blaschke products, especially their relationship with elliptic rational functions in filter design theory.

We give in this section a brief account of the theory of Jacobian elliptic functions. For more details we refer to [9]. For all $\tau \in \mathbb{H}$ we write $q = e^{\pi i \tau}$ with the branch of $q^{1/4}$ chosen such that $q^{1/4} = e^{-\pi/4}$ at $\tau = i$ and recall, following the notation of

Tannery-Molk, the four theta functions

$$\begin{array}{llll} \vartheta_1(v,\tau) = & \sum_{n=-\infty}^{\infty} i^{2n-1} q^{(n+\frac{1}{2})^2} e^{(2n+1)vi}, & \vartheta_2(v,\tau) & = \sum_{n=-\infty}^{\infty} q^{(n+\frac{1}{2})^2} e^{(2n+1)vi}, \\ \vartheta_3(v,\tau) = & \sum_{n=-\infty}^{\infty} q^{n^2} e^{2nvi}, & \vartheta_0(v,\tau) & = \sum_{n=-\infty}^{\infty} (-1)^n q^{n^2} e^{2nvi} \end{array}$$

as well as the following special functions

$$\begin{aligned} \omega_1 &= \omega_1(\tau) = \pi\, \vartheta_3^2(0,\tau) = \pi(1 + 2q + 2q^4 + \cdots)^2, \\ \omega_2 &= \omega_2(\tau) = \omega_1(\tau) \cdot \tau, \\ k &= k(\tau) = \vartheta_2^2(0,\tau)/\vartheta_3^2(0,\tau), \\ \sqrt{k} &= \sqrt{k}(\tau) = \vartheta_2(0,\tau)/\vartheta_3(0,\tau), \\ k' &= k'(\tau) = \vartheta_0^2(0,\tau)/\vartheta_3^2(0,\tau), \\ \sqrt{k'} &= \sqrt{k'}(\tau) = \vartheta_0(0,\tau)/\vartheta_3(0,\tau), \\ \lambda &= \lambda(\tau) = k^2(\tau) = \vartheta_2^4(0,\tau)/\vartheta_3^4(0,\tau). \end{aligned}$$

If $\tau \in \mathbb{H}$ is purely imaginary then $\sqrt{k}$ and ω_1 are both positive real numbers. We shall write simply $\vartheta_1(v)$ instead of $\vartheta_1(v,\tau)$ when no ambiguity arises and similar remark applies to many other functions. Following Jacobi [21, p.512] his elliptic functions can be defined by

$$\operatorname{sn} u = \frac{1}{\sqrt{k}} \cdot \frac{\vartheta_1(u/\omega_1)}{\vartheta_0(u/\omega_1)},\ \operatorname{cn} u = \frac{\sqrt{k'}}{\sqrt{k}} \cdot \frac{\vartheta_2(u/\omega_1)}{\vartheta_0(u/\omega_1)},\ \operatorname{dn} u = \sqrt{k'} \cdot \frac{\vartheta_3(u/\omega_1)}{\vartheta_0(u/\omega_1)}.$$

The elliptic function sn takes $2\omega_1$ and ω_2 as a pair of primitive periods and satisfies

$$\operatorname{sn}(\pm\omega_1/2) = \pm 1 \tag{10}$$

as well as

$$\operatorname{sn}(\omega_1 - u) \;=\; \operatorname{sn} u. \tag{11}$$

Moreover the critical points $\mathfrak{D}_{\mathrm{sn}} = \left\{\frac{\omega_1}{2}, \frac{3\omega_1}{2}, \frac{\omega_1+\omega_2}{2}, \frac{3\omega_1+\omega_2}{2}\right\} + \Lambda_{2\omega_1,\omega_2}$. In [21, p.145] Jacobi expressed his functions as infinite products

$$\vartheta_0(v) = c\prod_{n=1}^{\infty}(1 - q^{2n-1}e^{2\pi iv})(1 - q^{2n-1}e^{-2\pi iv}), \tag{12}$$

$$\vartheta_1(v) = c\,q^{1/4}\, 2\sin\pi v \prod_{n=1}^{\infty}(1 - q^{2n}e^{2\pi iv})(1 - q^{2n}e^{-2\pi iv}), \tag{13}$$

$$\operatorname{sn}(u) = \frac{1}{\sqrt{k}} \frac{q^{1/4}\, 2\sin\frac{\pi u}{\omega_1} \prod_{n=1}^{\infty}(1 - q^{2n}e^{2\pi i\frac{u}{\omega_1}})(1 - q^{2n}e^{-2\pi i\frac{u}{\omega_1}})}{\prod_{n=1}^{\infty}(1 - q^{2n-1}e^{2\pi i\frac{u}{\omega_1}})(1 - q^{2n-1}e^{-2\pi i\frac{u}{\omega_1}})} \tag{14}$$

where $c = \prod_{n=1}^{\infty}(1 - q^{2n})$. Glaisher introduced nine other elliptic functions in [19, p.86] among which $\operatorname{cd} u := \operatorname{cn} u/\operatorname{dn} u$ is of particular importance in the sequel. By the addition formula of sn we have

$$\operatorname{cd} u \;=\; \operatorname{sn}(u + \omega_1/2). \tag{15}$$

Since both cn and dn are even we have

$$\mathrm{cd}\, u \quad = \quad \mathrm{cd}(-u). \tag{16}$$

The elliptic function cd also takes $2\omega_1$ and ω_2 as a pair of primitive periods. It follows immediately from (16) that for $\tau \in \mathbb{H}$ the function cd is a special analytic representation of the Kummer map, namely $\mathrm{cd} : E_{2\omega_1,\omega_2} \to E_{2\omega_1,\omega_2}/\langle -1 \rangle \xrightarrow{\sim} \mathbb{P}^1$.

Later we shall make use of the nice relation

$$\gamma(t) \quad = \quad \sqrt{k\left(\frac{4ti}{\pi}\right)} \tag{17}$$

between γ introduced in Section 5 and the elliptic modulus k.

7 Elliptic rational functions

The concept of elliptic rational function is rarely found in the mathematical literature, but it is of central importance for advanced filter design. A nice treatment of elliptic rational functions in engineering can be found in [25, Chapter 12]. Here we shall consider more generally elliptic rational functions in a universal family $\mathcal{T}_{n,\tau}$ parameterized by $\tau \in \mathbb{H}$ to be constructed below. This will be more satisfactory in mathematics. In this section we shall work out that normalized Chebyshev-Blaschke products $f_{n,t}$ agree with the set $\{\mathcal{T}_{n,\tau} : \tau \in \mathbb{R}_+ i\}$.

For $\tau \in \mathbb{H}$ and $n \in \mathbb{N}$ there is a natural isogeny

$$[n] : E_{2\omega_1(\tau),\omega_2(\tau)} \to E_{2\omega_1(n\tau),\omega_2(n\tau)} \text{ given by } [n](z) = nz\omega_1(n\tau)/\omega_1(\tau)$$

which descends through the Kummer map to a rational function $\mathcal{T}_{n,\tau}$ and one obtains the commutative diagram.

$$\begin{array}{ccc} E_{2\omega_1(\tau),\omega_2(\tau)} & \overset{[n]}{\to} & E_{2\omega_1(n\tau),\omega_2(n\tau)} \\ \downarrow cd & & \downarrow cd \\ \mathbb{P}^1 & \overset{\mathcal{T}_{n,\tau}}{\to} & \mathbb{P}^1 \end{array}$$

The map given by the function cd is an analytic representation of the Kummer map. Obviously $z_1 \equiv \pm z_2 \mod \Lambda_{2\omega_1(\tau),\omega_2(\tau)}$ *implies that* $nz_1 \equiv \pm nz_2 \mod \Lambda_{2\omega_1(n\tau),\omega_2(n\tau)}$ and this shows that the map $[n]$ is invariant under the action given by the involution. By the theory of descent it induces therefore a rational map $\mathcal{T}_{n,\tau}$ as stated.

We call a rational function $f \in \mathbb{C}(z)$ *elliptic* if there exist a positive integer n, τ in $\mathbb{H}$ and ϵ, ε in $\mathrm{Aut}_{\mathbb{C}}(\mathbb{P}^1)$ such that $\epsilon \circ f \circ \varepsilon = \mathcal{T}_{n,\tau}$. The nesting property

$$\mathcal{T}_{m,n\tau} \circ \mathcal{T}_{n,\tau} = \mathcal{T}_{n,m\tau} \circ \mathcal{T}_{m,\tau}$$

easily follows from the construction and is very important for elliptic filter theory.

If f is a finite Blaschke product then $|f(z)| = 1$ for all $z \in \mathbb{T}^1$ and this follows from $|z - a| = |1 - \overline{a}z|$ for all $z \in \mathbb{T}^1$. By similar but more involved arguments we use Jacobi products and prove

Proposition 7.1. *If τ is a purely imaginary point of the upper half plane $\mathbb{H}$ and if there exists $m \in \mathbb{Z}$ such that $\frac{i\Im v}{\tau} = \frac{2m+1}{4}$ then*

$$|\vartheta_0(v)| = |\vartheta_1(v)|.$$

Proof. The elliptic function $\frac{\vartheta_1(v)}{\vartheta_0(v)}$ has primitive periods 2 and τ and therefore it suffices to prove the claim under the assumption $\frac{i\Im v}{\tau} = \frac{1}{4}$ or $\frac{i\Im v}{\tau} = \frac{3}{4}$. We shall only verify this in the case $\frac{i\Im v}{\tau} = \frac{1}{4}$ since similar arguments apply in the remaining case. By the product formulae (12) and (13) and by the trivial fact that $v = \Re v + \tau/4$ we have

$$\begin{aligned}
\vartheta_0(v) &= c\prod_{n=1}^{\infty}(1 - e^{(2n-\frac{1}{2})\pi i\tau}\, e^{2\pi i\Re v})\,(1 - e^{(2n-\frac{3}{2})\pi i\tau}\, e^{-2\pi i\Re v}),\\
\vartheta_1(v) &= c\, e^{\frac{\pi i\tau}{4}}\, 2\sin(\pi v)\prod_{n=1}^{\infty}(1 - e^{(2n+\frac{1}{2})\pi i\tau}\, e^{2\pi i\Re v})\,(1 - e^{(2n-\frac{1}{2})\pi i\tau}\, e^{-2\pi i\Re v})
\end{aligned}$$

and we have to show that both terms have the same absolute value. Our assumption $-i\tau > 0$ gives $e^{(2n\pm\frac{1}{2})\pi i\tau} \in \mathbb{R}$ and leads to

$$\begin{aligned}
\overline{1 - e^{(2n-\frac{1}{2})\pi i\tau}\, e^{2\pi i\Re v}} &= 1 - e^{(2n-\frac{1}{2})\pi i\tau}\, e^{-2\pi i\Re v},\\
\overline{1 - e^{(2n+\frac{1}{2})\pi i\tau}\, e^{2\pi i\Re v}} &= 1 - e^{(2n+\frac{1}{2})\pi i\tau}\, e^{-2\pi i\Re v}.
\end{aligned}$$

We use these two identities to compare the infinite products above and see that for the proof of the proposition it will be sufficient to verify that

$$|1 - e^{\frac{\pi i\tau}{2}}\, e^{-2\pi i\Re v}| = |2e^{\frac{\pi i\tau}{4}}\, \sin(\pi v)|.$$

This follows from

$$\begin{aligned}
|1 - e^{\frac{\pi i\tau}{2}}\, e^{-2\pi i\Re v}| &= |1 - e^{\frac{\pi i\tau}{2}}\, e^{2\pi i\Re v}|\\
&= |1 - e^{2\pi i v}|\\
&= |1 - \cos(2\pi v) - \sin(2\pi v)i|\\
&= |2\sin^2(\pi v) - 2\sin(\pi v)\cos(\pi v)i|\\
&= |2\sin(\pi v)e^{(\pi v - \frac{\pi}{2})i}|\\
&= |2\sin(\pi v)e^{-\pi\Im v}|\\
&= |2e^{\frac{\pi i\tau}{4}}\, \sin(\pi v)|
\end{aligned}$$

and completes the proof. □

Corollary 7.2. *Let τ be a purely imaginary point of the upper half plane $\mathbb{H}$ and m a rational integer. If $\frac{4m+1}{4} < \frac{i\Im v}{\tau} < \frac{4m+3}{4}$ then*

$$|\vartheta_0(v)| < |\vartheta_1(v)|$$

and if $\frac{4m-1}{4} < \frac{i\Im v}{\tau} < \frac{4m+1}{4}$ then

$$|\vartheta_0(v)| > |\vartheta_1(v)|.$$

Proof. The elliptic function $\varphi(v) = \frac{\vartheta_1(v)}{\vartheta_0(v)}$ is of order 2 and takes $2, \tau$ as a pair of primitive periods. We take the parallelogram with vertex $0, 2, 2+\tau, \tau$ as a fundamental domain. By Proposition 7.1 each of the images of $\left\{z : \frac{i\Im z}{\tau} = \frac{4m+1}{4}, m \in \mathbb{Z}\right\}$ and of $\left\{z : \frac{i\Im z}{\tau} = \frac{4m+3}{4}, m \in \mathbb{Z}\right\}$ under φ covers $\mathbb{T}$. Together with the fact that $\deg \varphi = 2$ this leads to $\varphi^{-1}(\mathbb{T}) = \left\{z : \frac{i\Im z}{\tau} = \frac{2m+1}{4}, m \in \mathbb{Z}\right\}$. If our second claim is not true then there exists w such that $-\frac{1}{4} < \frac{i\Im w}{\tau} < \frac{1}{4}$ and $|\vartheta_0(w)| \leq |\vartheta_1(w)|$. Moreover by $\varphi(0) = 0$ we have $|\vartheta_0(0)| \geq 0 = |\vartheta_1(0)|$ and by the continuity of $|\varphi(v)|$ there exists z such that $-\frac{1}{4} < \frac{i\Im z}{\tau} < \frac{1}{4}$ and $|\varphi(z)| = 1$. This contradicts our previous conclusion on $\varphi^{-1}(\mathbb{T})$ and proves our second claim. The first assertion is obtained in a similar way. □

Corollary 7.3. *If τ is a purely imaginary point in the upper half plane $\mathbb{H}$ then*

$$\begin{aligned}
\mathrm{sn}^{-1}\left\{z : |z| = \frac{1}{\sqrt{k}}\right\} &= \left\{w : \frac{i\Im w}{\omega_2} = \frac{2m+1}{4}, m \in \mathbb{Z}\right\}, \\
\mathrm{sn}^{-1}\left\{z : |z| < \frac{1}{\sqrt{k}}\right\} &= \left\{w : \frac{4m-1}{4} < \frac{i\Im w}{\omega_2} < \frac{4m+1}{4}, m \in \mathbb{Z}\right\}, \\
\mathrm{sn}^{-1}\left\{z : |z| < \frac{1}{\sqrt{k}}\right\} &= \left\{w : \frac{4m+1}{4} < \frac{i\Im w}{\omega_2} < \frac{4m+3}{4}, m \in \mathbb{Z}\right\}
\end{aligned}$$

and the same holds with sn *replaced by* cd.

Proposition 7.4. *If τ is a purely imaginary point in the upper half plane $\mathbb{H}$ then*

$$\mathrm{sn}^{-1}[-1,\, 1] = \{w : i\Im w = m\omega_2, m \in \mathbb{Z}\}$$

and the same holds with sn *replaced by* cd.

Proof. First of all we recall that as remarked in Section 6 the assumption $-i\tau > 0$ implies that q, $\sqrt{k}$ and ω_1 are all positive real numbers. If w is a real number then the quotient $v = w/\omega_1$ and

$$\mathrm{sn} w = \frac{1}{\sqrt{k}} \frac{2q^{1/4}\sin \pi v - 2q^{9/4}\sin 3\pi v + 2q^{25/4}\sin 5\pi v - \cdots}{1 - 2q\cos 2\pi v + 2q^4\cos 4\pi v - 2q^9\cos 6\pi v + \cdots}$$

are also real. The elliptic function sn takes $2\omega_1, \omega_2$ as a pair of primitive periods and the vertices $0, 2\omega_1, 2\omega_1+\omega_2, \omega_2$ define a fundamental domain. Furthermore we have seen in Section 6 that $\mathrm{sn}\frac{\omega_1}{2} = 1$ and $\mathrm{sn}\frac{3\omega_1}{2} = -1$ and that the critical points of sn within the fundamental domain are given by $\left\{\frac{\omega_1}{2}, \frac{3\omega_1}{2}, \frac{\omega_1+\omega_2}{2}, \frac{3\omega_1+\omega_2}{2}\right\}$. These facts imply that the image of $[0, 2\omega_1]$ under sn covers $[-1, 1]$ twice and we conclude that the preimage of $[-1, 1]$ in the fundamental domain by the twofold covering sn is $[0, 2\omega_1]$ which leads to the desired statement. □

Corollary 7.3 and 7.4 applied to the function $f(z) = \sqrt{k(n\tau)}\mathcal{T}_{n,\tau}(z/\sqrt{k(\tau)})$ gives

Proposition 7.5. *If $\tau \in \mathbb{H}$ is purely imaginary then f is a finite Blaschke product with $f(\sqrt{k(\tau)}) = \sqrt{k(n\tau)}$ and $f([-\sqrt{k(n\tau)}, \sqrt{k(n\tau)}]) = [-\sqrt{k(\tau)}, \sqrt{k(\tau)}]$.*

This together with Proposition 5.3 and (17) leads to

Corollary 7.6. *The Blaschke products $f_{n,t}(z)$ are elliptic with respect to $\tau = 4ti/\pi$; in other words we have*

$$f_{n,t}(z) = \sqrt{k(4nti/\pi)}\mathcal{T}_{n,4ti/\pi}(z/\sqrt{k(4ti/\pi)}).$$

8 Factorizations of finite Blaschke products

In this section we give a detailed study of the factorization properties of finite endomorphisms of the unit disk. If $f \notin \mathrm{Aut}_{\mathbb{C}}(\mathbb{E})$ is a finite endomorphism of the unit disk then the following two Propositions follow directly from Theorem 1.1 and Theorem 1.2.

Proposition 8.1. *The finite map f is composite if and only if its monodromy group is imprimitive.*

In the introduction we introduced the length of f with respect to a prime factorization as the number of its factors. As a corollary of Theorem 1.2 we have

Proposition 8.2. *The length of f is independent of prime factorizations.*

Proof. We choose a path α sufficiently close to $\mathbb{T}$ and apply Theorem 1.2 to get the assertion. □

Lemma 8.3. *If f and g are finite Blaschke products and if $z^n \circ g = f \circ z^n$ then f takes the form $f(z) = z^m h(z)^n$ where $m = ord_0 f$ and h is a finite Blaschke product.*

Proof. It suffices to prove that for any nonzero p in $\mathbb{E}$ we have $ord_p f \equiv 0 \mod n$. We denote by $p^{1/n}$ any nth root of p and using the functional equation we obtain

$$\mathrm{ord}_p f \equiv \mathrm{ord}_{p^{1/n}}(f \circ z^n) \equiv \mathrm{ord}_{p^{1/n}}(z^n \circ g) \equiv 0 \mod n$$

as desired. $\square$

Proof of Main Theorem 1.3. By Proposition 8.2 the length of f is independent of a given prime factorization. Moreover if

$$\mathbb{E} \xrightarrow{\varphi_1} \mathfrak{T}_1 \xrightarrow{\varphi_2} \mathfrak{T}_2 \rightarrow \cdots \rightarrow \mathfrak{T}_{r-1} \xrightarrow{\varphi_r} \mathbb{E}$$

is a decomposition of f into a product of finite maps then in particular for any $1 \leq i \leq r-1$ the map $\varphi_i \circ \varphi_{i-1} \circ \cdots \circ \varphi_1$ from $\mathbb{E}$ to $\mathfrak{T}_i$ is finite. This together with Lemma 2.3 implies that $\mathfrak{T}_i$ is biholomorphically equivalent to the unit disk. After taking finitely many operations of the first kind as described in the theorem our problem amounts to describe how one passes from one prime factorization

$$\mathbb{E} \xrightarrow{\varphi_1} \mathbb{E} \xrightarrow{\varphi_2} \mathbb{E} \rightarrow \cdots \rightarrow \mathbb{E} \xrightarrow{\varphi_r} \mathbb{E}$$

to another decomposition

$$\mathbb{E} \xrightarrow{\psi_1} \mathbb{E} \xrightarrow{\psi_2} \mathbb{E} \rightarrow \cdots \rightarrow \mathbb{E} \xrightarrow{\psi_r} \mathbb{E}$$

with all Riemann surfaces being unit disks. Furthermore by Example 2.2 (Fatou) all φ_i and all ψ_i are finite Blaschke products.

Let $\mathfrak{d}_f \subset \mathbb{E}$ be the set of critical values of f, $n = \deg f$ and $\mathfrak{L}$ the lattice of groups lying between $\pi_1(\mathbb{E} \setminus \mathfrak{d}_f)$ and $\pi_1(\mathbb{E} \setminus f^{-1}(\mathfrak{d}_f))$. If we write $G_i = \pi_1(\mathbb{E} \setminus (\varphi_r \circ \cdots \circ \varphi_i)^{-1}(\mathfrak{d}_f))$ and $K_i = \pi_1(\mathbb{E} \setminus (\psi_r \circ \cdots \circ \psi_i)^{-1}(\mathfrak{d}_f))$ then we have $G_1 = K_1 = \pi_1(\mathbb{E} \setminus f^{-1}(\mathfrak{d}_f))$ as well as $G_{r+1} = K_{r+1} = \pi_1(\mathbb{E} \setminus \mathfrak{d}_f)$ and by Corollary 2.5 applied to $\Sigma = \mathfrak{d}_f$, $q \notin \Sigma$ some point in $\mathbb{E}$ and p a point in $\mathbb{E}$ with $f(p) = q$ we deduce that our prime decompositions of f induce maximal chains

$$G_1 \leq G_2 \leq \cdots \leq G_r \leq G_{r+1}$$

and

$$K_1 \leq K_2 \leq \cdots \leq K_r \leq K_{r+1}$$

with G_i, K_i in $\mathfrak{L}$. We apply Theorem 1.2 to $\mathfrak{M} = \mathfrak{N} = \mathbb{E}$ and f and therefore we know from the proof of Theorem 1.2 that $\mathfrak{L}$ is a sublattice of $\mathfrak{L}_n$ which is in particular modular. By Proposition 3.2 we may pass inductively from the first chain to the second with only one change at each step. This gives a topological description of our algorithm using fundamental groups. Corollary 2.5 allows us to write down the algorithm in terms of explicit analytic maps as listed in the theorem. As explained at the end of Section 3 this boils down to solving the functional equation

$$\alpha_2 \circ \alpha_1 = h = \beta_2 \circ \beta_1 \tag{18}$$

with α_i, β_i prime Blaschke products, $\deg \alpha_1 = \deg \beta_2 = l$, $\deg \alpha_2 = \deg \beta_1 = k$ and $\gcd(k,l) = 1$. Our strategy is to get first a polynomial solution to this equation and then, using the monodromy representations given by such a solution, to transform the polynomial solution into a solution expressed in terms of Blaschke products.

Proposition 4.3 applied to (18) for some homeomorphism $i_0 = j_0 : \mathbb{E} \to \mathbb{C}$, which induces other homeomorphisms $i_1, j_1, i_2 = j_2$ from $\mathbb{E}$ to $\mathbb{C}$, leads to the following easily verified identities,

$$(i_2, i_0)_* h = (i_1, i_0)_* \alpha_2 \circ (i_2, i_1)_* \alpha_1 = (j_1, i_0)_* \beta_2 \circ (i_2, j_1)_* \beta_1$$

and hence we have a solution to the two polynomial equation

$$\alpha_2 \circ \alpha_1 = h = \beta_2 \circ \beta_1$$

with α_i, β_i prime polynomials, $\deg \alpha_1 = \deg \beta_2 = l, \deg \alpha_2 = \deg \beta_1 = k$ and $\gcd(k,l) = 1$. The polynomial solutions to this equation can be written out by Ritt's work [35]. Accordingly there exist linear polynomials ι_i such that one of the identities

1) $\iota_1 \circ (i_2, i_1)_* \alpha_1 \circ \iota_2 = \iota_3 \circ (j_1, i_0)_* \beta_2 \circ \iota_4 = z^l$;

2) $\iota_1 \circ (i_1, i_0)_* \alpha_2 \circ \iota_2 = \iota_3 \circ (i_2, j_1)_* \beta_1 \circ \iota_4 = z^k$;

3) $\iota_1 \circ (i_2, i_1)_* h \circ \iota_2 = T_{lk}$.

is satisfied. In case 1) of the list above α_1 and β_2 are totally ramified maps from $\mathbb{E}$ to $\mathbb{E}$. After finitely many operations of the first kind we may assume that $\alpha_1 = \beta_2 = z^l$. Then the functional equation (18) reduces to $\alpha_2 \circ z^l = z^l \circ \beta_1$ and Lemma 8.3 gives the solution as desired. Similar considerations apply to case 2).

If we are in case 3) the monodromy of h is a Chebyshev representation and therefore h is a Chebyshev-Blaschke product as explained in Lemma 5.1. After another finitely many operations of the first kind we may assume that $h, \alpha_1, \alpha_2, \beta_1, \beta_2$ are all normalized Chebyshev-Blaschke products and we are done. $\square$

9 Polydisks

In this section we sketch how to extend our results to the case of polydisks. Firstly we recall from Example 2.2 and Rischel's version [34] of Remmert-Stein's theorem [33] the following famous classification result.

Theorem 9.1 (Fatou-Remmert-Stein-Rischel)**.** *If f is an analytic map from $\mathbb{E}^d$ to $\mathbb{E}^d$ then it is finite if and only if*

$$f(z_1, \cdots, z_d) = (f_1(z_{\sigma(1)}), \cdots, f_n(z_{\sigma(d)}))$$

with $\sigma \in S_d$ and f_j finite Blaschke products.

This theorem together with the results proved in Section 8 shows that *if f is a nonlinear finite map from $\mathbb{E}^d$ to $\mathbb{E}^d$ then it is composite if and only if its monodromy group is imprimitive.* In addition *the length of a nonlinear finite map $f : \mathbb{E}^d \to \mathbb{E}^d$ is independent of prime factorizations* and one sees that this leads without any difficulty to a higher dimensional generalization of our main theorem

Theorem 9.2. *Given two prime factorizations of a nonlinear finite map $f : \mathbb{E}^d \to \mathbb{E}^d$, one can pass from one to the other by repeatedly uses of explicit operations.*

Acknowledgements: Our figures are made by Mimi Lui, the short proof of Lemma 2.3 is suggested by Ngaiming Mok and the functorial formulation of Section 4 is suggested by Gisbert Wüstholz. We thank them for their kindly help. The first author would like to thank Gisbert Wüstholz for his invitation and the hospitality of the Forschungsinstitut für Mathematik(FIM) and Department of Mathematics, ETH-Zürich. Lastly we would like to thank the referee for his careful reading and many helpful suggestions.

References

[1] L.V. Ahlfors, Finitely generated Kleinian groups, Amer. J. Math. **86** (1964), 413–429.

[2] L.V. Ahlfors, L. Sario, *Riemann surfaces,* Princeton Mathematical Series, No. **26** Princeton University Press, Princeton, N.J. 1960.

[3] N.L. Alling, The valuation theory of meromorphic function fields. *Entire Functions and Related Parts of Analysis.* Amer. Math. Soc., Providence, R.I., 1968, 8–29.

[4] R.M. Avanzi, U. Zannier, Genus one curves defined by separated variable polynomials and a polynomial Pell equation, Acta Arith. **99** (2001), 227–256.

[5] A.F. Beardon, T.W. Ng. *On Ritt's factorization of polynomials.* J. London Math. Soc. **2** 62 (2000), no. 1, 127–138.

[6] Yu.F. Bilu, R.F. Tichy, The Diophantine equation $f(x) = g(y)$, Acta Arith. **95** (2000), no. 3, 261–288.

[7] T.S. Blyth. *Lattices and ordered algebraic structures.* Springer-Verlag London, Ltd., London, 2005.

[8] N. Bourbaki, *General Topology. Chapter 1-4.* Springer-Verlag, Berlin, 1998.

[9] K. Chandrasekharan, *Elliptic functions,* Springer-Verlag, 1985.

[10] P.J. Cameron. *Permutation groups.* Cambridge University Press, Cambridge,1999.

[11] T.C. Dinh, Ensembles d'unicité pour les polynômes, Ergodic Theory Dynam. Systems **22** (2002), no. 1, 171–186.

[12] F. Dorey, G. Whaples. Prime and composite polynomials, J. Algebra **28** (1974), 88–101.

[13] H.T. Engstrom. Polynomial substitutions, Amer. J. Math. **63**, (1941). 249–255.

[14] M. Fried, On a theorem of Ritt and related Diophantine problems, J. Reine Angew. Math. **264** (1973), 40–55.

[15] P. Fatou, Sur les équations fonctionelles, Bull. Soc. Math. France. **48** (1920), 208–314.

[16] P. Fatou, Sur les fonctions holomorphes et bornées à l'intérieur d'uncercle, Bull. Soc. Math. France. **51** (1923), 191–202.

[17] O. Forster. *Lectures on Riemann surfaces.* Springer-Verlag, New York, 1991.

[18] D. Ghioca, T.J. Tucker, M.E. Zieve, Intersections of polynomials orbits, and a dynamical Mordell-Lang conjecture, Invent. Math. **171** (2008), no. 2, 463–483.

[19] J.W.L. Glaisher, On elliptic functions, Messenger of Mathematics, **6** (1882), 81–95.

[20] A. Grothendieck, Letter to Faltings, (1983).

[21] C.G.J. Jacobi, *Gesammelte Werke*, Zweite Ausgabe Chelsea Publishing Co., New York, 1969.

[22] J. Klaus. *Topology.* Springer, New York, 1984.

[23] H. Levi, Composite polynomials with coefficients in an arbitrary field of characteristic zero, Amer. J. Math. **64**, (1942), 389–400.

[24] S.K. Lando, A.K. Zvonkin, *Graphs on surfaces and their applications.* Springer-Verlag, Berlin, 2004.

[25] M.D. Lutovac, D.V. Tosic, B.L. Evans, *Filter Design for Signal Processing using MATLAB and Mathematica.* New Jersey, USA, Prentice Hall, 2001.

[26] P. Müller, *Primitive monodromy groups of polynomials. Recent developments in the inverse Galois problem (Seattle, WA, 1993)*, 385–401, Contemp. Math., **186**, Amer. Math. Soc., Providence, RI, 1995.

[27] P. Müller, M.E. Zieve, On Ritt's polynomial decomposition theorems, arXiv:0807.3578.

[28] M. Muzychuk, F. Pakovich, Jordan-Hölder theorem for imprimitivity systems and maximal decompositions of rational functions, arXiv:0712.3869.

[29] F. Pakovich, On polynomials sharing preimages of compact sets, and related questions. Geom. Funct. Anal. **18** (2008), no. 1, 163–183.

[30] F. Pakovich, *Prime and composite Laurent polynomials*, Bull. Sci. Math. **133** (2009), no. 7, 693–732.

[31] T. Radó, Zur Theorie der mehrdeutigen konformen Abbildungen. (German) Acta Litt. Sci. Szeged. **1**, 55-64 (1922).

[32] R. Remmert, *Classical topics in complex function theory.* Springer-Verlag, New York, 1998.

[33] R. Remmert, K. Stein, Eigentliche holomorphe Abbildungen, Math.Z. **73** (1960), 159–189.

[34] H. Rischel, Ein Satz über eigentliche holomorphe Abbildungen von analytischen Polyedergebieten, Math.Scand. **14** (1964), 220–224.

[35] J.F. Ritt, Prime and composite polynomials, Trans. Amer. Math. Soc. **23** (1922), no. 1, 51–66.

[36] J.S. Robinson, *A course in the theory of groups.* Springer-Verlag, New York, 1996.

[37] I.R. Shafarevich, *Algebraic geometry I.* Algebraic curves. Algebraic manifolds and schemes. Springer-Verlag, Berlin, 1994.

[38] H. Völklein, *Groups as Galois groups: An introduction.* Cambridge University Press, Cambridge, 1996.

[39] M.X. Wang, http://hub.hku.hk/handle/123456789/51854, Thesis, (2008).

[40] U. Zannier, Ritt's second theorem in arbitrary characteristic. J. Reine Angew. Math. **445** (1993), 175–203.