Annales Academiæ Scientiarum Fennicæ
Series A. I. Mathematica
Volumen 18, 1993, 125–146

GRAPHS AND FUNCTIONAL EQUATIONS

Lowell J. Hansen and Harold S. Shapiro

Wayne State University, Department of Mathematics
Detroit, MI 48202, U.S.A.; lowell@math.wayne.edu
Royal Institute of Technology, Department of Mathematics
S-10044 Stockholm, Sweden; shapiro@math.kth.se

Abstract. Given two polynomials P and Q and a class of functions (e.g., meromorphic, entire, rational, polynomial), we study conditions under which there exist no nonconstant functions f and g in the given class which satisfy the functional equation $f(P(z)) = g(Q(z))$.

1. Introduction

Let P be entire and h meromorphic on the complex plane, $\mathbf{C}$. For each $w \in \mathbf{C}$, the *w-level set* of P is the set $\{z : P(z) = w\}$. We will say that h *is a function of* P if h is constant on the w-level set of P for every w. It is easy to see that this is equivalent to the existence of a meromorphic function f on $\mathbf{C}$ such that $h = f \circ P$.

This paper and a sequel in preparation deal with the following problem. Let k denote a given meromorphic function on $\mathbf{C}$, and P, Q given *polynomials*. Look for all pairs of meromorphic functions f, g on $\mathbf{C}$ such that

$$f\bigl(P(z)\bigr) - g\bigl(Q(z)\bigr) = k(z). \tag{1}$$

As we shall see in the sequel, (1) is closely related to Dirichlet's problem for the Laplace equation, especially the question of the harmonic continuation of the solution in the exterior of the domain where the problem is posed.

In this paper we consider the special case where $k = 0$. Then (1) becomes

$$f\bigl(P(z)\bigr) = g\bigl(Q(z)\bigr) =: h(z), \tag{2}$$

in which case we say that h *is a function of both* P *and* Q. Thus we could phrase our goal as follows: Given the polynomials P and Q and a class of functions (meromorphic, entire or polynomials), we wish to determine all nonconstant functions h of the given class which are functions of both P and Q. Clearly, for a smaller class of allowable solutions, it is less likely that nonconstant solutions will exist. For many pairs (P, Q) it will turn out that (2) has no nonconstant solution.

1991 Mathematics Subject Classification: Primary 30D05; Secondary 39B32.

This will in turn imply that on certain algebraic curves no nonconstant functions of the given class can have vanishing real part, results of a type studied earlier in [FNS]. Special cases of our investigation are a theorem of Rényi and Rényi [RR] and its generalization by Fuchs and Gross [FG] concerning periodicity of functions of the form $(f \circ P)$.

Another way of formulating the problem is as follows. Let $\mathscr{M} = \mathscr{M}(\mathbf{C})$, $\mathscr{E} = \mathscr{E}(\mathbf{C})$ denote the meromorphic and entire functions on the complex plane, respectively. For a given polynomial P, $\mathscr{M}_P$ and $\mathscr{E}_P$ denote, respectively, the meromorphic and entire functions that are "functions of P". Then the question of whether, for given polynomials P and Q, any nonconstant function h satisfies (2) is the question of the existence of nonconstant functions in $\mathscr{M}_P \cap \mathscr{M}_Q$.

An important tool in studying these questions is the notion of cyclic elements in the group of invertible germs of analytic functions. This is classical but, as it is not easy to find accessible references, we recall this material in Section 2. In order to have a handy language for some of the considerations in this paper, we introduce in Section 3 an infinite graph $\Gamma(P,Q)$ induced by any given pair of polynomials P, Q of degrees ≥ 2. Its nodes, or vertices, are the points of $\mathbf{C}$. Two vertices z, ζ of $\Gamma(P,Q)$ are connected by an edge if and only if at least one of the relations $P(z) = P(\zeta)$, $Q(z) = Q(\zeta)$ holds. It is then clear that h satisfying (2) takes the same value at two points whenever they can be joined by a path (finite succession of edges) of this graph. Thus, for example, if infinitely many distinct points lying in a bounded region can be joined by paths to one and the same point, h must be constant.

The balance of the paper is organized as follows. In Section 4 we present the history of the problem. For expository purposes, we recap the original proofs of a theorem of Rényi and Rényi [RR] and its generalization by Fuchs and Gross [FG]. In Section 5 we link together the notions of Sections 2 and 3 to provide an alternative mechanism for studying the problem and giving computer-assisted proofs that certain functional equations (2) have no nonconstant meromorphic solutions. We conclude in Section 6 with various further remarks.

2. The group of invertible germs

By a *germ* we shall mean a function φ analytic on some neighborhood of 0 in the complex plane satisfying $\varphi(0) = 0$. We say φ is *invertible* if $\varphi'(0) \neq 0$. In this case, the equation $w - \varphi(z) = 0$ is uniquely solvable for z (for w sufficiently near 0) and we get $z = \psi(w)$ for some germ ψ. We will also denote ψ by $\varphi^{[-1]}$, and call it the *inverse* of φ. It is easy to see that, under functional composition, the invertible germs form a group, which we shall denote by $\mathscr{G}$. The identity of $\mathscr{G}$ is, of course, the germ e defined by $e(z) \equiv z$.

We summarize the main information we need in the following

Proposition 2.1. *The following are equivalent for an element φ of $\mathscr{G}$ and integer $k \geq 1$.*

(i) *$\varphi^{[k]}$ is the identity of $\mathscr{G}$, where $\varphi^{[k]}$ denotes $\varphi \circ \varphi \cdots \circ \varphi$ (k times).*
(ii) *There exists $\psi \in \mathscr{G}$ and a k^{th} root of unity (denoted by ω) such that*

$$\varphi(z) = \psi^{[-1]}\big(\omega\psi(z)\big)$$

(in other words, φ is conjugate to the rotation $\zeta \mapsto \omega\zeta$).
(iii) *There exists a function f holomorphic near 0, having a Taylor expansion of the form $z^k +$ (*higher order terms*), such that $f \circ \varphi = f$.*

Proof. Assume first that (iii) holds. Then, f can be written as g^k where g is holomorphic near 0 and has the expansion

$$g(z) = z + (\text{higher order terms}).$$

Then (iii) implies that

$$g\big(\varphi(z)\big)^k = g(z)^k$$

so there is some k^{th} root ω of 1 such that

$$g\big(\varphi(z)\big) = \omega g(z).$$

This yields (ii), with $\psi = g$.

Now assume (ii) holds. Then

$$\varphi^{[2]}(z) = \varphi\big(\varphi(z)\big) = \psi^{[-1]}\big(\omega\psi\big(\varphi(z)\big)\big)\psi^{[-1]}\Big(\omega\psi\big(\psi^{[-1]}\big(\omega\psi(z)\big)\big)\Big) = \psi^{[-1]}\big(\omega^2\psi(z)\big)$$

Proceeding iteratively, we conclude that

$$\varphi^{[k]}(z) = \psi^{[-1]}\big(\omega^k\psi(z)\big) = z,$$

proving (i).

Finally, we will show that (i) implies (iii). Assume (i) holds, and let

$$f(z) := z\varphi(z)\varphi^{[2]}(z)\cdots\varphi^{[k-1]}(z).$$

Clearly f is holomorphic near $z = 0$ and its power series has the form

$$z^k + (\text{higher order terms}).$$

Moreover, $f\big(\varphi(z)\big) = f(z)$, verifying (iii). □

Remark. An element φ of $\mathscr{G}$ satisfying (i) for some k is said to be *cyclic.* The *order* of φ is then the smallest k such that $\varphi^{[k]}$ is the identity. It is easy to see that φ is cyclic of order k if and only if (ii) holds for some $\psi \in \mathscr{G}$ and *primitive* k^{th} root ω of unity.

3. Reentrant graphs

Our basic question is closely linked with a purely combinatorial one: Let P, Q be polynomials with complex coefficients, of degrees p, q, respectively. We assume $p \geq 2$ and $q \geq 2$. We define a graph (with infinitely many nodes) $\Gamma(P,Q)$ as follows: The nodes are the points of the complex plane, $\mathbf{C}$. Given two points z_1, z_2 $(z_1 \neq z_2)$ of $\mathbf{C}$, we connect them by an edge if and only if at least one of the relations $P(z_1) = P(z_2)$, $Q(z_1) = Q(z_2)$ holds. Moreover, we "color" the edge "red" if the first relation holds, "blue" if the second holds, and "red/blue" if both hold. This colored graph is denoted $\Gamma(P,Q)$.

A *path* on $\Gamma(P,Q)$ is a sequence of directed edges such that the initial point of each edge after the first is the endpoint of the preceding. The path is *proper* if the edges comprising it alternate in color. A red/blue edge may be considered as red, or blue as desired; thus, the sequences (red, red/blue, red, red/blue) and (red/blue, red/blue) alternate. A *closed* proper path has, of course, an even number of edges. (It is easy to see that if we were to allow, say, two or more consecutive "red" edges, they could always be replaced by just one, so the requirement that paths be proper is not really restrictive.) We will call a closed proper path consisting of m edges an *m-cycle*.

Clearly, if $h \in \mathscr{M}_P \cap \mathscr{M}_Q$, then $h(z) = h(z_1)$ for each z reachable from z_1 by a path of finite length along $\Gamma(P,Q)$. (An alternative formalism, in terms of equivalence relations is given at the end of this section.)

We will say that $\Gamma(P,Q)$ *has the clustering property* if, within some bounded region D, there are infinitely many distinct points connected to some fixed point z_1 along $\Gamma(P,Q)$. Then, obviously, if $\Gamma(P,Q)$ *has the clustering property, each function in $\mathscr{M}_P \cap \mathscr{M}_Q$ is constant.*

$\Gamma(P,Q)$ *has the weak clustering property* if there is a compact set $K \subset \mathbf{C}$ such that for every positive integer N, there is a point connectable to N distinct points of K.

Proposition 3.1. *If $\Gamma(P,Q)$ is weak clustering, then $\mathscr{M}_P \cap \mathscr{M}_Q$ consists only of constants.*

Proof. Suppose not, and $h \in \mathscr{M}_P \cap \mathscr{M}_Q$ is nonconstant meromorphic. Then, by assumption, for each N we have $z_1, z_2, \ldots, z_N$ in K such that

$$h(z_1) = h(z_2) = \cdots = h(z_N) = w_N.$$

Then, either $w_N \to \infty$ or else $\{w_N\}$ has a bounded subsequence. Clearly, it is sufficient to treat the second case, since otherwise we consider $1/h$. So, by passing to a subsequence if necessary, we may assume $w_n \to \lambda \in \mathbf{C}$. Let now γ be a contour surrounding K, on which $h(z)$ does not take the value λ. Then, for large N, h does not take the value w_N on γ. Now

$$J_N := \frac{1}{2\pi i} \int_\gamma \frac{h'(z)}{h(z) - w_N}\, dz$$

is the number of roots of $h(z) = w_N$ inside γ, minus the number of poles of h. As $N \to \infty$, the former is at least N, so $J_N \to +\infty$. However,

$$\lim_{N\to\infty} J_N = \frac{1}{2\pi i}\int_\gamma \frac{h'(z)}{h(z)-\lambda}\,dz$$

which is finite. □

We define the graph $\Gamma(P,Q)$ to be *reentrant* if there exists a compact set $K \subset \mathbf{C}$ such that every $\zeta \in \mathbf{C}$ can be connected to some point in K by a path in $\Gamma(P,Q)$.

A relationship between reentrant graphs and our basic question is set forth in the following theorem.

Theorem 3.1. *If $\Gamma(P,Q)$ is reentrant and h_1 and h_2 belong to $\mathscr{M}_P \cap \mathscr{M}_Q$, then there exists a nontrivial polynomial π with $\pi(h_1,h_2) = 0$.*

Proof. The hypothesis that h_1 and h_2 belong to $\mathscr{M}_P \cap \mathscr{M}_Q$ implies that there exist f_i, g_i meromorphic on $\mathbf{C}$ with

$$\left\{\begin{array}{l} h_1(z) = f_1\big(P(z)\big) = g_1\big(Q(z)\big), \\ h_2(z) = f_2\big(P(z)\big) = g_2\big(Q(z)\big). \end{array}\right. \tag{3}$$

By a well-known procedure (cf. [AS, p. 52]) there exists a nontrivial polynomial π such that the meromorphic function $h = \pi(h_1,h_2)$ has no poles in $\big\{|z| \le R\big\}$. Also, because of equation (3), h_1 and h_2 (and hence h) are constant along the vertices of each proper path in $\Gamma(P,Q)$. Therefore, all values taken by h in $\mathbf{C}$ are taken in $\big\{|z| \le R\big\}$, which implies that h is a bounded entire function. Hence h is constant. □

Corollary 3.1. *If h is any entire function satisfying* (2) *for suitable entire f, g and $\Gamma(P,Q)$ is reentrant, then h is constant.*

We will use the notation $n(z_*,h)$ to denote the *multiplicity* at z_* of the nonconstant meromorphic function h. That is, $n(z_*,h)$ is the order of the zero of $h - h(z_*)$ at z_* if h is analytic at z_*, and is the order of the pole at z_* otherwise.

Remarks. 1. The multiplicity of a nonconstant meromorphic function differs from 1 only on the isolated set where it has a pole of order larger than 1 or where its derivative is zero. Thus, the multiplicity of a nonconstant meromorphic function is bounded on compact sets.

2. Let P and f be nonconstant meromorphic functions with P analytic at z_* and f analytic at $w = P(z_*)$. Then there exists p_* analytic at z_* with $p_*(z_*) \ne 0$ so that

$$P(z) - w = (z - z_*)^{n(z_*,P)} p_*(z).$$

Using a similar formula for f near w and taking compositions, it follows that

$$n(z_*, f \circ P) = n(z_*,P)\,n\big(P(z_*),f\big).$$

The same formula also holds when f has a pole at $P(z_*)$.

Proposition 3.2. *If $\Gamma(P,Q)$ is reentrant (with associated compact set K), there is, to each nonconstant $h \in \mathscr{M}_P \cap \mathscr{M}_Q$, an integer $N = N_h$ such that*
(a) *all poles of h have multiplicity $\leq N$*
(b) *for each $a \in \mathbf{C}$, all roots of $h(z) = a$ have multiplicity $\leq N$.*

Proof. Pick any point z in $\mathbf{C}$ for which $h(z) = a$. There is a path $(z, z_1, \ldots, z_n, \zeta)$ connecting it to a point $\zeta \in K$.

Suppose first that P' and Q' are nonzero at all vertices of this path. If, say $P(z) = P(z_1)$, then it follows from Remark 2 above that the multiplicities of the a-points which h has at z and z_1 are the same. Continuing along the chain to ζ we see that the multiplicity in question is that of the a-point at ζ. But this is bounded independently of a by Remark 1 above.

In case, as we move along the path, we reach a point where P', say, vanishes to order r, then the multiplicity of the a-point can be multiplied by $r+1$ in the passage. Taking account that there are at most $(\deg P) - 1$ such points (and similarly for Q), we see that the multiplicity of h at ζ cannot exceed that at z by more than a factor C depending only on the degrees of P and Q. □

Theorem 3.1 demonstrates that it is of interest to have conditions which imply that the graph $\Gamma(P,Q)$ induced by a pair of polynomials P and Q is reentrant. The next theorem gives such a condition:

Theorem 3.2. *Let P and Q be monic polynomials of degrees p and q, respectively. Suppose that the rational function*

$$R(z) = \frac{P(z)^q}{Q(z)^p}$$

is nonconstant and that its Laurent expansion about ∞ has the form

$$R(z) = 1 + az^{-s} + \mathscr{O}\big(|z|^{-s-1}\big) \tag{4}$$

where $s \geq 1$ and $a \neq 0$. If

(5) *neither p nor q divides s and at least one of them does not divide $2s$,*

then $\Gamma(P,Q)$ is reentrant.

Remarks. 1. Before giving the proof, we give some examples that illustrate the theorem.

(a) Let $P(z) = (z-1)^3$ and $Q(z) = z^3$. Here, $R(z) = (1 - z^{-1})^9 = 1 - 9z^{-1} + \cdots$, so $s = 1$ and the hypotheses of the theorem are satisfied. In this case the conclusion can be verified directly, since it is easy to see that the vertices of $\Gamma(P,Q)$ reachable from ζ are exactly the orbit of ζ under the group generated by the maps $z \mapsto (2-z)$ and $z \mapsto \omega z$, where $\omega = e^{2\pi i/3}$.

(b) If P and Q are monic of the same degree $p \geq 3$ and $(P - Q)$ is neither constant nor of degree $p/2$, the conditions of our theorem are again satisfied. Indeed, $(P/Q) - 1 = (P - Q)/Q$ is, for large z, asymptotic to cz^{k-p} where $k = \deg(P-Q)$ and $c \neq 0$. Thus $P/Q = 1+cz^{k-p}+\cdots$, $(P/Q)^p = 1+pcz^{k-p}+\cdots$ and so the s in equation (4) is $(p-k)$ and (5) holds by virtue of the conditions on $(P - Q)$. To see why $(P - Q)$ may not be of degree $p/2$, consider the following example: Let $P(z) = z^4 + 2z^2 + 1$ and $Q(z) = z^4$. Then P is $P_1(z^2)$ and Q is $Q_1(z^2)$ (with $P_1(z) = (z+1)^2$ and $Q_1(z) = z^2$). Now there exists a nonconstant entire function, f, with $f(P_1) = f(Q_1)$. And so, composing with z^2, we get a nonconstant entire solution to $f(P) = f(Q)$. Notice that here $p = 4$ and the degree of $(P - Q)$ is $p/2 = 2$.

(c) If $P = T_p$ and $Q = T_q$, where T_p and T_q are Chebyshev polynomials of the first kind, the hypotheses of the theorem do not hold. Verifying this is simple if we replace $T_p(z)$ and $T_q(z)$ by the monic polynomials $2T_p(z/2)$ and $2T_q(z/2)$, respectively, which we now relabel as P and Q, respectively. To check condition (4), we suppose that $p < q$ and seek first to determine the degree of

$$P(z)^q - Q(z)^p = \left(2\cos\left(p\cos^{-1}(z/2)\right)\right)^q - \left(2\cos\left(q\cos^{-1}(z/2)\right)\right)^p .$$

If we replace $\cos^{-1}(z/2)$ by t and express $\cos(pt)$ and $\cos(qt)$ in terms of complex exponentials, the degree we seek is the largest r such that a u^r term appears in

$$(u^p + u^{-p})^q - (u^q + u^{-q})^p.$$

Now u^{pq} cancels and the next term in the expansions is a constant times $u^{p(q-2)}$ (since $p < q$). Thus $(P - Q)$ is of degree $p(q - 2)$, the s of (4) is $2p$ and so condition (5) is not satisfied. Also, since $T_p \circ T_q = T_q \circ T_p = T_{pq}$, the points of the graph $\Gamma(P, Q)$ to which a given ζ can be connected is simply the finite set where T_{pq} takes on the same value that it does at ζ, and so the graph $\Gamma(P, Q)$ is not reentrant.

2. If d denotes the greatest common divisor of (p, q), then R in (4) may be replaced by

$$\frac{P(z)^{(q/d)}}{Q(z)^{(p/d)}}$$

with no resulting change in the value of s.

3. Note that condition (5) implies that $\max(p, q) \geq 3$.

Proof of Theorem 3.1. The first step in the proof is contained in:

Proposition 3.3. *Given P and Q satisfying the hypotheses of the theorem, there exist positive numbers M, t with $t < 1$ such that for $|\zeta| > M$, one of the following holds:*

(i) *there exists z satisfying $P(z) = P(\zeta)$ and $\left|Q(z)\right| \leq t|Q(\zeta)|$; or,*

(ii) *there exists z satisfying $Q(z) = Q(\zeta)$ and $\left|P(z)\right| \leq t\left|P(\zeta)\right|$.*

Suppose that this proposition is proved. The theorem may then be deduced as follows: Let $K = \{z : |z| \le M\} \cup \{z : |P(z)Q(z)| \le 1\}$. Then K is compact. Now, start with any ζ, $|\zeta| > M$ and apply the proposition. Suppose that (i) holds, and let $P(z_1) = P(\zeta)$, $|Q(z_1)| \le t|Q(\zeta)|$. Then $|P(z_1)Q(z_1)| \le t|P(\zeta)Q(\zeta)|$. (Note that this last inequality would have followed similarly in the case that (ii) holds.) We continue inductively as follows: Suppose that z_j has been produced with $|P(z_j)Q(z_j)| \le t^j|P(\zeta)Q(\zeta)|$. If $|z_j| \le M$, then z_j belongs to K and we are done. If $|z_j| > M$, apply the proposition again to find z_{j+1} with $|P(z_{j+1})Q(z_{j+1})| \le t|P(z_j)Q(z_j)|$. The inductive hypothesis then implies that $|P(z_{j+1})Q(z_{j+1})| \le t^{j+1}|P(\zeta)Q(\zeta)|$. In this way we produce a path in $\Gamma(P,Q)$ for which we either eventually reach a z_j with $|z_j| \le M$ or else a z_j with $|P(z_j)Q(z_j)| \le 1$ by virtue of the inequality $|P(z_j)Q(z_j)| \le t^j|P(\zeta)Q(\zeta)|$. Consequently, in either case, we enter K after finitely many steps. Thus, the proof of the theorem will be complete once this proposition is proved.

Proof of Proposition 3.3. In the proof of this proposition, we will need the following lemma:

Lemma 3.1. *Let P be monic of degree p. Then there exist constants C_1, C_2 such that for all $\zeta \in \mathbf{C}$, $|\zeta| \ge C_1$, the roots of $P(z) = P(\zeta)$ distinct from ζ satisfy*

$$|z_j - \omega_j\zeta| \le C_2; \qquad j = 1, 2, \dots, p-1 \tag{6}$$

where $\omega_j = e^{2\pi ij/p}$.

Proof. Let $\{z_j\}_{j=1}^{p-1}$ denote the roots of

$$P(z) = P(\zeta) \tag{7}$$

other than ζ. We will show that for $j = 1, 2, \dots, p-1$, $z_j = \omega_j\zeta + \mathscr{O}(1)$.

First, fix j and $a > 0$ and let $D_j(a)$ denote the disk $\{|z - \omega_j\zeta| < a|\zeta|\}$. Easy estimates show that there exist positive constants C_1 and a so that if $|\zeta| > C_1$,

$$\left|(z^p - \zeta^p) - \big(P(z) - P(\zeta)\big)\right| < |z^p - \zeta^p|$$

for z on $\partial D_j(a)$. Then, by Rouché's theorem, $\big(P(z) - P(\zeta)\big)$ has inside this disk as many zeroes as $(z^p - \zeta^p)$, namely 1.

To get more precise information, consider (for fixed $j, 1 \le j \le p-1$) the polynomial

$$k(t) = P(\omega_j\zeta + t) - P(\zeta).$$

Thus k has p roots. By our first estimate, it is easy to see that there are $p-1$ roots $t_1, \dots, t_{p-1}$ (corresponding to $(\omega_j\zeta + t) \sim \omega_k\zeta$ with $k \ne j$) satisfying $|t_i| \ge a|\zeta|$

for some $a > 0$ depending only on P. Thus, the product of these roots of k is, in absolute value, $\geq \bigl(a|\zeta|\bigr)^{p-1}$. But, the product of absolute values of *all* the roots of k is $\bigl|P(\omega_j\zeta) - P(\zeta)\bigr|$, which is $\leq A|\zeta|^{p-1}$. We conclude that the other root of k has absolute value $\leq A/a^{p-1}$, which we now call C_2. This completes the proof of the lemma since we have shown that to each j there is a root of $P(z) = P(\zeta)$ with $|z - \omega_j\zeta| \leq C_2$. □

Remark. One can produce finer asymptotics of the roots as follows: Again, let

$$k(t) := P(\omega_j\zeta + t) - P(\zeta) = \bigl[P(\omega_j\zeta) + tP'(\omega_j\zeta) + \cdots + t^p\bigr] - P(\zeta).$$

If t_0 denotes the root of k with $|t_0| \leq C_2$, this gives

$$\bigl|P(\omega_j\zeta) - P(\zeta) + t_0 P'(\omega_j\zeta)\bigr| \leq C_3\bigl(1 + |\zeta|\bigr)^{p-2}.$$

Hence,

$$t_0 = \frac{P(\zeta) - P(\omega_j\zeta)}{P'(\omega_j\zeta)} + \mathscr{O}\bigl(|\zeta|^{-1}\bigr).$$

This estimate will be needed later.

To complete the proof of the proposition, we first try, with z_j as in the lemma, to find j with

$$P(z_j) = P(\zeta) \qquad \text{and} \qquad \bigl|Q(z_j)\bigr| \leq t\bigl|Q(\zeta)\bigr| \tag{8}$$

for some $t < 1$. We assume $M \geq C_1$, so that (6) holds. Now,

$$\Bigl|\frac{Q(\zeta)}{Q(z_j)}\Bigr|^p = \Bigl|\frac{P(z_j)^q}{Q(z_j)^p}\Bigr| \cdot \Bigl|\frac{Q(\zeta)^p}{P(\zeta)^q}\Bigr| = \bigl|R(z_j)\bigr| \cdot \bigl|R(\zeta)\bigr|^{-1}. \tag{9}$$

Also, from (4), we have, putting $a = Ae^{i\alpha}$, $A > 0$

$$\bigl|R(re^{i\theta})\bigr|^2 = 1 + 2Ar^{-s}\cos(s\theta - \alpha) + \mathscr{O}(r^{-s-1}) \tag{10}$$

for large r. Writing $\zeta = \varrho e^{i\varphi}$, $z_j = r_j e^{i\theta_j}$, we get from (9) and (10):

$$\begin{aligned}\Bigl|\frac{Q(\zeta)}{Q(z_j)}\Bigr|^{2p} &= \bigl(1 + 2Ar_j^{-s}\cos(s\theta_j - \alpha) + \mathscr{O}(r_j^{-s-1})\bigr)\cdot \\ &\quad \cdot \bigl(1 + 2A\varrho^{-s}\cos(s\varphi - \alpha) + \mathscr{O}(\varrho^{-s-1})\bigr)^{-1} \\ &= 1 + 2Ar_j^{-s}\cos(s\theta_j - \alpha) - 2A\varrho^{-s}\cos(s\varphi - \alpha) + \mathscr{O}(\varrho^{-s-1})\end{aligned} \tag{11}$$

Now, by (6),

$$\big|r_j e^{i\theta_j} - \varrho e^{i(\varphi+(2\pi j/p))}\big| \le C_2; \qquad j = 1, 2, \ldots, p-1. \tag{12}$$

Also, $|r_j - \varrho| \le C_2$, so $|1 - r_j/\varrho| \le C_2\varrho^{-1}$ and (12) implies

$$\big|e^{i\theta_j} - e^{i(\varphi+(2\pi j/p))}\big| \le C_3\varrho^{-1}$$

(where, henceforth all C_j denote constants that can depend only on P and Q). Consequently,

$$\big|\theta_j - \big(\varphi + (2\pi j/p)\big)\big| \le C_4\varrho^{-1}. \tag{13}$$

From these estimates it is clear that if we replace r_j and θ_j by ϱ and $\varphi + 2\pi j/p$, respectively, in the right side of (12) we only introduce an error that is $\mathscr{O}(\varrho^{-s-1})$. Hence

$$\Big|\frac{Q(\zeta)}{Q(z_j)}\Big|^{2p} = 1 + 2A\varrho^{-s}\left[\cos\big(s\big(\varphi + (2\pi j/p)\big) - \alpha\big) - \cos(s\varphi - \alpha)\right] + \mathscr{O}(\varrho^{-s-1})$$

for $j = 1, 2, \ldots, p-1$. Suppose now there is a constant $\lambda(p,q) > 0$ such that

$$\max_{1\le j\le (p-1)} \left[\cos\big(s\big(\varphi + (2\pi j/p)\big) - \alpha\big) - \cos(s\varphi - \alpha)\right] \ge \lambda(p,q) \tag{14}$$

(where $\lambda(p,q)$ does not depend on φ, α). Then, for some $j \in \{1, 2, \ldots, p-1\}$

$$\Big|\frac{Q(\zeta)}{Q(z_j)}\Big|^{2p} \ge 1 + 2A\lambda(p,q)\varrho^{-s} + \mathscr{O}(\varrho^{-s-1}).$$

This leads easily to

$$\big|Q(z_j)\big| \le \left[1 - \delta\varrho^{-s} + \mathscr{O}(\varrho^{-s-1})\right] \cdot \big|Q(\zeta)\big|$$

for some $\delta > 0$ depending only on p, q, P, Q. Thus,

$$\big|Q(z_j)\big| \le [1 - \delta\varrho^{-s}] \cdot \big|Q(\zeta)\big| \tag{15}$$

holds for some $j \in \{1, 2, \ldots, p-1\}$, if $|\zeta| = \varrho \ge C_5$. We thus have: *If* (14) *holds, then for* $|\zeta| \ge C_5$, *we can find* z *with* $P(z) = P(\zeta)$ *and* $\big|Q(z)\big| \le t\big|Q(\zeta)\big|$ *for some* $t < 1$. *This* $t = t(\zeta)$ *is uniformly less than 1 on bounded* ζ*-sets.*

Let us examine condition (14). If it fails, there is an example of this with $\alpha = 0$, as we see by putting $\varphi - (\alpha/s) = \varphi'$. Now, the function

$$F(\varphi) = \max_{1\le j\le (p-1)} \left[\cos s\big(\varphi + (2\pi j/p)\big) - \cos s\varphi\right]$$

is continuous, of period 2π, and so if it is *positive* for each φ, it is bounded below by a positive constant $\lambda(p,q)$. This can only fail if all the inequalities

$$\cos s\bigl(\varphi + (2\pi j/p)\bigr) - \cos s\varphi \le 0, \qquad j = 1, 2, \ldots, p-1 \tag{16}$$

hold. In exactly the same way we can show that, for $|\zeta| \ge C_6$, unless all the inequalities

$$\cos s\bigl(\varphi + (2\pi k/q)\bigr) - \cos s\varphi \ge 0, \qquad k = 1, 2, \ldots, q-1 \tag{17}$$

hold, there is a point z with $Q(z) = Q(\zeta)$ and $\bigl|P(z)\bigr| \le t\bigl|P(\zeta)\bigr|$ (here again, $t < 1$ and is bounded away from 1 on bounded ζ-sets).

Thus, the proposition, and the theorem, will be demonstrated once we show: *the inequalities* (16) *and* (17) *cannot all hold.* Indeed, we will show that a contradiction follows from the assumption that they all hold. Summing (16) from $j = 1$ to $p-1$, and (17) from $k = 1$ to $q-1$ gives $-p\cos s\varphi \le 0$ and $-q\cos s\varphi \ge 0$, respectively (here it is crucial that $e^{2\pi i s/p}$ and $e^{2\pi i s/q}$ are both different from 1). Hence $\cos s\varphi = 0$ and (16), (17) become

$$-\sin(2\pi js/p)\sin s\varphi \le 0, \qquad j = 1, 2, \ldots, p-1 \tag{18}$$

and

$$-\sin(2\pi ks/q)\sin s\varphi \ge 0, \qquad k = 1, 2, \ldots, q-1. \tag{19}$$

But at least one of p, q (say, p) does not divide $2s$. Then (18) cannot hold, because $\sin s\varphi = \pm 1$, and the numbers $\sin(2\pi s/p)$ and $\sin\bigl(2\pi(p-1)s/p\bigr)$ are nonzero and negatives of one another. Thus, either the term with $j = 1$ or that with $j = p-1$ is *positive* in (18). This completes the proof of Proposition 3.3 and, thereby, the proof of Theorem 3.1. ▫

Before closing this section, we remark on a more general formulation of the basic combinatorial problem underlying $\Gamma(P,Q)$. Given two equivalence relations, $\equiv_1$ and $\equiv_2$, on a set (like $\mathbf{C}$), there is induced a new equivalence relation $\equiv_{1,2}$ as the "smallest" equivalence relation $\sim$ such that $x \equiv_1 y$ or $x \equiv_2 y$ implies that $x \sim y$. Since the relation "$x \equiv_1 y$ or $x \equiv_2 y$" is not transitive, we must define $x \sim y$ so that it holds if and only if there is a finite sequence $z_1, \ldots, z_n$ such that "$x \equiv_1 z_1$ or $x \equiv_2 z_1$", and "$z_1 \equiv_1 z_2$ or $z_1 \equiv_2 z_2$", and $\cdots$ "$z_n \equiv_1 y$ or $z_n \equiv_2 y$" all hold. Thus, the equivalence classes are precisely sets of vertices of the graph gotten by joining two points whenever they are either 1-equivalent, or 2-equivalent, that are connected with one another, that is components of $\Gamma(P,Q)$. In particular, if the underlying set is $\mathbf{C}$ and we define $z_1 \equiv_P z_2$ to mean that $P(z_1) = P(z_2)$, then $\equiv_{P,Q}$ is the equivalence relation corresponding to components of $\Gamma(P,Q)$. Thus, we can reformulate, for example, the definition of *weak clustering*, as: *To each compact $K \subset \mathbf{C}$ and N there is a set of N distinct points in K which are $\equiv_{P,Q}$ equivalent.*

4. History of the problem

In one sense, our problem originates in work of F. Marty ([M1] and [M2]) and T. Shimizu ([Shi1], [Shi2] and [Shi3]) from the 1930's dealing with the non-Möbius (and even, non single-valued) automorphisms of analytic functions. These questions were later taken up and given a systematic exposition by Gunnar af Hällström in several papers of which we especially mention [H1] and [H2]. Paper [H1] contains numerous references to the work of Marty and Shimizu. In [H1, p. 13], af Hällström writes (we translate from the German): "Here the question arises whether, given (rational) functions $f_1, \dots, f_m$ a rational (nonconstant) function g can be found which is simultaneously a rational function of each of the f_j." This is essentially the problem studied in the present paper, except that

(i) the question as posed by af Hällström is *more general* in that m functions are allowed rather than 2, and these may be *rational*, not merely polynomials, while

(ii) his question is also *less general* in that g is required to be *rational*, not merely meromorphic.

However, apart from noting some trivial examples where no g exists, he does not go further into the problem. In Section 6 below we shall present a method that yields nontrivial necessary conditions for the existence of a (nonconstant) rational g.

Another case of our problem turned up in the work of Rényi and Rényi [RR], who proved: $\mathscr{E}_P \cap \mathscr{E}_{\mathrm{per}\,1}$ is trivial when $\deg P \geq 3$. (Here, $\mathscr{E}_{\mathrm{per}\,1}$ is the set of entire functions of period 1.)

Their proof is very easy. Suppose that $f \in \mathscr{E}_P$ has period 1. Look at a large disk $D = \{z : |z| < R\}$ and let $z_0 \in \partial D$ be a point where $|f(z_0)| = \max_{z \in D} |f(z)|$. By periodicity f takes the same value $f(z_0)$ at all points $z_0 +$ integer. If any of these enters D, then f is constant by the maximum modulus theorem. If none of them enters D, then clearly z_0 is close to either the north pole or south pole of ∂D. But if $\deg P = p \geq 3$ and R is fairly large, then Lemma 3.1 shows that f takes the value $f(z_0)$ at points close to $(e^{2\pi i j/p}) \cdot z_0$, $j = 1, 2, \dots, p-1$. Since $p \geq 3$, at least one of these numbers has imaginary part of absolute value much less than R, so $f(z_0) = f(z_1)$ for some z_1 whose integer translates really enter D. (In terms of earlier terminology: the graph induced by P and the "periodicity 1" partition is reentrant.)

Fuchs and Gross [FG] showed that a suitably modified theorem holds also for meromorphic functions, and the proof is very similar. Before turning to this, however, let us look at some examples: Let $P(z) = (z-a)^p$, $Q(z) = (z-b)^q$ where p, q are at least as large as 2 and $a \neq b$. $h \in \mathscr{M}_P$ means $h(z_1) = h(z_2)$ whenever $(z_1 - a)^p = (z_2 - a)^p$. That is, $z_1 - a = \omega(z_2 - a)$ for ω some p^{th} root of unity. In other words, $\mathscr{M}_P$ is the set of meromorphic functions invariant under the group generated by the substitution

$$z \mapsto \omega z + (1-\omega)a, \qquad \omega = e^{2\pi i/p}. \tag{20}$$

Likewise, $\mathscr{M}_Q$ is the set of meromorphic functions invariant under

$$z \mapsto \lambda z + (1-\lambda)b, \qquad \lambda = e^{2\pi i/q}. \tag{21}$$

It is now fairly easy to decide when $\Gamma(P,Q)$ has any of the basic properties embodied in earlier definitions. For example, it is clustering if $p = q \geq 7$, it is reentrant (but not weak clustering) for $p = q = 3$. There are a finite number of choices of $p \geq 2$, $q \geq 2$ such that (20), (21) generate a properly discontinuous group; in these cases, and only these, $\mathscr{M}_P \cap \mathscr{M}_Q$ is nontrivial.

What Fuchs and Gross really prove is: *If $p = \deg P$ and $\cos 2\pi/p$ is irrational, then the graph Γ generated by P and "periodicity 1" is clustering.* Here is an outline of the proof. We will use $\equiv$ to denote the equivalence relation generated by $\equiv_P$ and $\equiv_{\mathrm{per}\,1}$ (notation as introduced in the last paragraph of Section 3). We will assume without loss of generality that $P(z) = z^p + cz^{p-k} + \cdots$ with $k \geq 2$. This can be accomplished, if necessary, by replacing P by $P \circ \varphi$, where φ is a linear polynomial. Then, if N is a large integer, the remark following Lemma 3.1 implies that the roots of $P(z) = P(N)$ are $\omega^j N + \mathscr{O}(N^{-1})$, $j = 1, 2, \ldots, p$.

Therefore,

$$\begin{aligned} 0 &\equiv N \equiv N\omega^2 + \mathscr{O}(N^{-1}) \equiv N\omega^2 + N + \mathscr{O}(N^{-1}) \\ &\equiv \overline{\omega}(N\omega^2 + N) + \mathscr{O}(N^{-1}) + M \equiv 2N \cos 2\pi/p + M + \mathscr{O}(N^{-1}) \end{aligned}$$

where M is any integer. Assume here that $\omega^2 + 1 \neq 0$, so that $N\omega^2 + N$ is large when N is. If now $\cos 2\pi/p$ is irrational, then we can, for each $\xi \in \mathbf{R}$ find a sequence $N_j \to \infty$, $M_j \to \infty$ such that $2N_j \cos 2\pi/p + M_j \to \xi$. Thus, 0 is $\equiv$ to a sequence which converges to ξ. Now, $\cos 2\pi/p$ is irrational unless $p = 1, 2, 3, 4, 6$. To see this, note that, since $\omega + \overline{\omega} = 2\cos 2\pi/p$, ω satisfies the quadratic equation $z^2 - 2(\cos 2\pi/p)z + 1 = 0$. So, if $\cos 2\pi/p = r/s$, where r, s are relatively prime integers, we see that s must be 1 or 2 since otherwise ω is not an algebraic integer. Thus $\cos 2\pi/p = 0$, ± 1, $\pm 1/2$ corresponding to $p = 4, 1, 2, 3, 6$, respectively. So, for $p = \deg P$ not equal to 1, 2, 3, 4, or 6, $\mathscr{M}_P$ contains no nontrivial periodic function.

Fuchs and Gross also find the exceptional cases: If $\cos 2\pi/p$ is rational, then we have, for each z:

$$\begin{aligned} z &\equiv z + n \text{ (for all integers } n) \equiv \omega^2(z+n) + \mathscr{O}(n^{-1}) \\ &\equiv \omega^2(z+n) + n + \mathscr{O}(n^{-1}) \equiv \omega^2 z + (\omega^2+1)n + \mathscr{O}(n^{-1}) \\ &\equiv \omega z + (\omega + \overline{\omega})n + \mathscr{O}(n^{-1}). \end{aligned}$$

If $\omega^2+1=0$ (that is, $p=4$), then letting n approach infinity we get a sequence (all terms of which are equivalent to z) converging to $\omega^2 z$ and thus $h(z)=h(\omega^2 z)$. That is, for $p=4$, $h(z)=h(-z)$. If, on the other hand, $\omega^2+1\neq 0$, then, since $(\omega+\overline{\omega})$ is rational, there are arbitrarily large n such that $n(\omega+\overline{\omega})$ is an integer. Letting $n\to\infty$ through such values, we get a sequence converging to ωz all of which are equivalent to z and thus $h(z)=h(\omega z)$. The complete theorem is gotten by a few more similar steps and we will not give them here. It leads to the choices $P(z)=(z-a)^p$ for arbitrary a.

In [FNS] algebraic curves in the plane were exhibited with the property that the real part of nonconstant meromorphic functions on $\mathbf{C}$ cannot vanish on them (examples are the curves given in Cartesian coordinates as the graphs of $y=x^n$, where $n\geq 3$). Actually, [FNS] worked with entire rather than meromorphic functions, but the main technique employed, based on invertible germs, applies to the meromorphic problem without changes. The relation of the problem studied in [FNS] to that of this paper can be seen as follows. Suppose, for example, that f is nonconstant and meromorphic on $\mathbf{C}$, and $\operatorname{Re}\big(f(z)\big)=0$ on the curve $\Gamma:=\big\{(x,x^3):x\in\mathbf{R}\big\}$. Then $\operatorname{Re}\big(f(x+ix^3)\big)=0$, $x\in\mathbf{R}$ so, denoting the meromorphic function $\overline{f(\overline{z})}$ by $f^*(z)$, we have $f(x+ix^3)=-f^*(x-ix^3)$, $x\in\mathbf{R}$. By analytic continuation we have then

$$f(z+iz^3)=g(z-iz^3),\qquad z\in\mathbf{C}$$

where $g=-f^*$, and this is a relation of type (2) with $P(z)=z+iz^3$, $Q(z)=z-iz^3$. It is important to observe that the failure of curves such as Γ above to be level curves of harmonic functions is a global, not a local feature: There are indeed functions holomorphic (and nonconstant) on a neighborhood of Γ, with real part vanishing on Γ. It is only when we try to make this neighborhood sufficiently large (e.g., the whole plane) that the task becomes impossible if we try to have at worst polar singularities.

L. Flatto [Fl] extended this work on level curves by giving a complete description of some exceptional cases arising in [FNS].

Finally, the paper [Sha] adapted some of the ideas from [FNS] to the study of the functional equation (2).

5. Eliminants

Let P, Q be polynomials with complex coefficients, of degrees p, q, respectively, and let $\Gamma(P,Q)$ denote the corresponding colored graph introduced in Section 3. Suppose now that z_1, z_2 are the endpoints of a red edge (that is, $P(z_1)=P(z_2)$) and assume that neither z_1 nor z_2 is a critical point of P (that is, both $P'(z_1)$ and $P'(z_2)$ are nonzero). Consider the function $u(s,t)=P(z_2+t)-P(z_1+s)$. It is holomorphic on $\mathbf{C}^2$ and $\partial u/\partial s$, $\partial u/\partial t$ are both nonzero at $(0,0)$. By the implicit function theorem there is a function $\varphi(s)$ holomorphic on a neighborhood of 0 such that $\varphi(0) = 0$ and $u\bigl(s, \varphi(s)\bigr) = 0$. Observe that, since $P\bigl(z_2 + \varphi(s)\bigr) - P(z_1 + s) = 0$, we have

$$\varphi'(0) = \frac{P'(z_1)}{P'(z_2)}.$$

Thus, φ is an element of the group $\mathscr{G}$ of invertible germs at 0 considered in Section 2. A similar consideration holds, of course, for "blue" edges. Thus, to each directed edge of $\Gamma(P, Q)$ whose endpoints are not critical points of P or Q is associated an element φ of $\mathscr{G}$.

Suppose now there exist nonconstant meromorphic functions f, g and h such that

$$h(z) = f\bigl(P(z)\bigr) = g\bigl(Q(z)\bigr). \tag{22}$$

Let Φ denote a proper closed path in $\Gamma(P, Q)$ with successive vertices $(z_1, z_2, \ldots, z_m, z_1)$ where $m \geq 2$ is an even integer. Suppose none of these vertices is a critical point of P or Q. We shall also suppose that no z_j is a pole of h. Let φ_1 denote the element of $\mathscr{G}$ associated to the edge (z_1, z_2) by the above procedure, φ_2 that associated to (z_2, z_3), etc. up to φ_m. Thus, we have the relations

$$\begin{aligned}
P(z_1) &= P(z_2), & P(z_1 + s) &= P\bigl(z_2 + \varphi_1(s)\bigr)\\
Q(z_2) &= Q(z_3), & Q(z_2 + s) &= Q\bigl(z_3 + \varphi_2(s)\bigr)\\
P(z_3) &= P(z_4), & P(z_3 + s) &= P\bigl(z_4 + \varphi_3(s)\bigr)\\
&\vdots\\
Q(z_m) &= Q(z_1), & Q(z_m + s) &= Q\bigl(z_1 + \varphi_m(s)\bigr).
\end{aligned}$$

By virtue of (22) we get from these relations

$$\begin{aligned}
h(z_1 + s) &= f\bigl(P(z_1 + s)\bigr) = f\bigl(P\bigl(z_2 + \varphi_1(s)\bigr)\bigr) = h\bigl(z_2 + \varphi_1(s)\bigr)\\
h(z_2 + s) &= g\bigl(Q(z_2 + s)\bigr) = g\bigl(Q\bigl(z_3 + \varphi_2(s)\bigr)\bigr) = h\bigl(z_3 + \varphi_2(s)\bigr)\\
&\vdots\\
h(z_m + s) &= g\bigl(Q(z_m + s)\bigr) = g\bigl(Q\bigl(z_1 + \varphi_m(s)\bigr)\bigr) = h\bigl(z_1 + \varphi_m(s)\bigr)
\end{aligned}$$

and so

$$\begin{aligned}
h(z_1 + s) = h\bigl(z_2 + \varphi_1(s)\bigr) = h\bigl(z_3 + \varphi_2\bigl(\varphi_1(s)\bigr)\bigr) = \cdots\\
= h\bigl(z_1 + \varphi_m\bigl(\varphi_{m-1}\bigl(\cdots \varphi_2(\varphi_1(s))\bigr)\bigr)\bigr).
\end{aligned}$$

That is, for the germ $\psi = (\varphi_m \circ \varphi_{m-1} \cdots \circ \varphi_1) \in \mathscr{G}$ we have

$$h(z_1 + s) = h\bigl(z_1 + \psi(s)\bigr). \tag{23}$$

We call ψ *the germ associated to the closed path* Φ.

Thus, in view of the discussion in Section 2, we can assert: Under the assumptions just made, ψ in (23) *is, for some* $n \geq 1$, *an* n^{th} *root of the identity in the group* $\mathscr{G}$, *and in particular* $\psi'(0)^n = 1$. It is not hard to see this is true also if the points z_j are poles of h.

Using this, one can sometimes show that, for a given pair of polynomials P, Q there are no nonconstant meromorphic solutions f, g to (22). Essentially this was done in [FNS] for the pair $P(z) = z + iz^m$, $Q(z) = z - iz^m$ by finding a closed proper path in $\Gamma(P,Q)$ for which the associated ψ was not a cyclic element of $\mathscr{G}$ (in fact $\psi'(0)$ was not a root of unity). We illustrate this technique for $P(z) = z^3 + z$, $Q(z) = z^3 - z$. To find z_1 and z_2 with $P(z_1) = P(z_2)$ and $z_1 \neq z_2$, we seek solutions to the equation $P^\#(z_1, z_2) = 0$, where

$$P^\#(z,w) = \frac{P(z) - P(w)}{z - w}. \tag{24}$$

Similarly, to find z_2 and z_3 with $Q(z_2) = Q(z_3)$ and $z_2 \neq z_3$, we solve the equation $Q^\#(z_2, z_3) = 0$, where

$$Q^\#(z,w) = \frac{Q(z) - Q(w)}{z - w}. \tag{25}$$

To find a 2-cycle z_1, z_2, z_1, the equations

$$\begin{aligned} P^\#(z_1, z_2) &= z_1^2 + z_1 z_2 + z_2^2 + 1 = 0 \\ Q^\#(z_2, z_1) &= z_1^2 + z_1 z_2 + z_2^2 - 1 = 0 \end{aligned}$$

would have to be satisfied. Clearly, this system has no solutions. For a 4-cycle, we need distinct points z_1, z_2, z_3, z_4 so that $P(z_1) = P(z_2)$, $Q(z_2) = Q(z_3), P(z_3) = P(z_4)$ and $Q(z_4) = Q(z_1)$. This translates into the system

$$\begin{aligned} z_1^2 + z_1 z_2 + z_2^2 + 1 &= 0 \\ z_2^2 + z_2 z_3 + z_3^2 - 1 &= 0 \\ z_3^2 + z_3 z_4 + z_4^2 + 1 &= 0 \\ z_4^2 + z_4 z_1 + z_1^2 - 1 &= 0. \end{aligned}$$

It can be seen that this implies that $z_1^4 + 1 = 0$. Beginning with the choice of $z_1 = \zeta = e^{\pi i/4}$, the above system then gives rise to the 4-cycle $z_1 = \zeta$, $z_2 = \zeta^3$, $z_3 = \zeta^5$, and $z_4 = \zeta^7$. Then, there must exist germs φ_1, φ_2, φ_3, φ_4 satisfying

$$\begin{aligned} P(z_1 + s) &= P\big(z_2 + \varphi_1(s)\big) \\ Q(z_2 + s) &= Q\big(z_3 + \varphi_2(s)\big) \\ P(z_3 + s) &= P\big(z_4 + \varphi_3(s)\big) \\ Q(z_4 + s) &= Q\big(z_1 + \varphi_4(s)\big). \end{aligned}$$

Hence if h, f, g are nonconstant meromorphic solutions to (22), the above chain produces $\psi = \varphi_4 \circ \varphi_3 \circ \varphi_2 \circ \varphi_1$ with $h(z_1 + s) = h\big(z_1 + \varphi(s)\big)$. Then there must exist a positive integer m (the *valence* of h at z_1) so that $\varphi^{[m]}$ is the identity map and $\big[\psi'(0)\big]^m = 1$. Now

$$\psi'(0) = \varphi_4'(0)\varphi_3'(0)\varphi_2'(0)\varphi_1'(0) = \frac{Q'(z_4)}{Q'(z_1)}\frac{P'(z_3)}{P'(z_4)}\frac{Q'(z_2)}{Q'(z_3)}\frac{P'(z_1)}{P'(z_2)}.$$

Then, from the definition of z_1, z_2, z_3, z_4 we get

$$\begin{aligned}\xi = \psi'(0) &= \frac{3(\zeta^7)^2 - 1}{3\zeta^2 - 1}\frac{3(\zeta^5)^2 + 1}{3(\zeta^7)^2 + 1}\frac{3(\zeta^3)^2 - 1}{3(\zeta^5)^2 - 1}\frac{3\zeta^2 + 1}{3(\zeta^3)^2 + 1} \\ &= \left(\frac{3\zeta^2 + 1}{3\zeta^2 - 1}\right)^4 = \left(\frac{3i + 1}{3i - 1}\right)^4 = \frac{-527}{625} - \frac{336}{625}i.\end{aligned}$$

But then,

$$\xi^2 + \frac{1054}{625}\xi + 1 = 0,$$

and so ξ (although it *is* of modulus 1) is not even an algebraic integer and so is not a root of unity. Consequently there can exist no nonconstant meromorphic functions f and g for which $f \circ P = g \circ Q$.

In case, for some proper closed path, the associated ψ turns out to be cyclic, we can nevertheless draw a strong conclusion:

Theorem 5.1. *Suppose that $\Gamma(P, Q)$ contains a proper closed path of length m (no vertices of which are singular points of P or Q) for which the associated ψ is cyclic of order n. Then, to each point $z \in \mathbf{C}$ there exists a proper closed path of length mn in $\Gamma(P, Q)$ having z as a vertex.*

Proof. The points of $\mathbf{C}$ reachable by a "red" edge starting at z_1 are the roots w of the equation $P^{\#}(z_1, w) = 0$. Starting from w, the points reachable by a "blue" edge are the roots ζ of $Q^{\#}(w, \zeta) = 0$ (here, $P^{\#}$ and $Q^{\#}$ are as defined in (24) and (25), respectively). Continuing in this fashion we arrive at the following proposition:

The points w reachable from z by a proper path of length $2k$ are the roots w of the equation $E_k(z, w) = 0$, where E_k (which we shall call the k^{th} *eliminant* of the ordered pair (P, Q)) is the polynomial obtained by eliminating $t_1, t_2, \ldots, t_{2k-1}$

from the equations

$$\begin{aligned} P^{\#}(z,t_1) &= 0 \\ Q^{\#}(t_1,t_2) &= 0 \\ P^{\#}(t_2,t_3) &= 0 \\ Q^{\#}(t_3,t_4) &= 0 \\ &\vdots \\ P^{\#}(t_{2k-2},t_{2k-1}) &= 0 \\ Q^{\#}(t_{2k-1},w) &= 0. \end{aligned}$$

Returning to the proof of the theorem, suppose we have a proper closed path $\Psi \subset \Gamma(P,Q)$ of length $m = 2k$ and vertices $(z_1, z_2, \ldots z_{2k}, z_1)$ with associated germ ψ which is an n^{th} root of the identity. For a point $\zeta = z_1 + t$ with $|t|$ sufficiently small, the calculations made earlier show that the points

$$z_1 + t,\ z_2 + \varphi_1(t),\ z_3 + \varphi_2\big(\varphi_1(t)\big), \ldots, z_1 + \varphi_{2k}\big(\cdots\big(\varphi_1(t)\big)\cdots\big)$$

are the vertices of a proper path starting at ζ. The last point is $z_1 + \psi(t)$ where $\psi = \varphi_{2k} \circ \cdots \circ \varphi_1$ is, by assumption, an n^{th} root of the identity in $\mathscr{G}$. Iterating, we see that $z_1 + \psi^{[r]}(t)$ is the endpoint of a proper path of length $2kr$ starting at ζ (recall that $\psi^{[r]}$ denotes the r^{th} iterate of ψ under composition). For $r = n$, $\psi^{[r]}$ is the identity, so the endpoint is ζ. That is, the path closes.

Consider now the eliminant $E_{kn}(z,w)$. For fixed z, its roots are all points w attainable as the endpoints of proper paths of length $2kn$ starting at z. The above discussion implies that $E_{kn}(\zeta,\zeta)$ vanishes for all ζ on a neighborhood of z_1, and hence identically. But this implies that, for each $\zeta \in \mathbf{C}$, there exists a proper path of length $2kn$ starting at ζ and terminating at ζ. □

This theorem allows one to give computer-assisted proofs that certain functional equations (22) have no nonconstant meromorphic solutions. As an example, consider the case $P(z) = z^3$, $Q(z) = z^3 + 3z^2 + \frac{9}{2}z$. To find a 2-cycle z_1, z_2, z_1, the equations

$$\begin{aligned} z_1^2 + z_1 z_2 + z_2^2 &= 0 \\ z_1^2 + z_1 z_2 + z_2^2 + 3(z_1 + z_2) + \frac{9}{2} &= 0 \end{aligned}$$

must be satisfied. This occurs for $z_1 = 3\omega/2$ and $z_2 = 3\omega^2/2$, where $\omega = e^{2\pi i/3}$. If ψ is the germ associated with this 2-cycle and $\xi = \psi'(0)$, a calculation shows that $\xi = -1$. Since this was a 2-cycle and -1 is a second root of unity, we conclude by Theorem 5.1 that if $h = f \circ P = g \circ Q$ admits a nonconstant solution, then every point in the plane lies on a 4-cycle. However, the corresponding eliminant $E_4(z,w)$, when w is replaced by z, does not vanish identically. Indeed, it equals

$$15116544(4z^2 + 6z + 9)^3(16z^4 + 24z^3 + 72z^2 - 54z + 81).$$

Consequently, nearly all points in the plane do not lie on a 4-cycle in $\Gamma(P,Q)$ and thus $h = f \circ P = g \circ Q$ possesses no nonconstant meromorphic solution.

Remark. We have seen that, when Φ is a proper closed path in $\Gamma(P,Q)$ with vertices $(z_1, z_2, \ldots, z_m, z_1)$ and none of the z_j is a zero of P' or Q', there is associated with Φ a certain $\psi \in \mathscr{G}$ whose *cyclicity* is necessary for the existence of nontrivial meromorphic f, g satisfying (2). It is easy to see that *if ψ is noncyclic, $\Gamma(P,Q)$ has the weak clustering property.* It is also trivial that $\Gamma(P,Q)$ is clustering if $\left|\psi'(0)\right| \neq 1$. By virtue of the theorem of C.L. Siegel [Si], a sufficient condition for strong clustering, in case $\left|\psi'(0)\right| = 1$, is (writing $\psi'(0) = e^{2\pi i t}$, $t \in \mathbf{R}$) that *there exist positive numbers a, b such that $\left|t - (m/n)\right| > an^{-b}$ holds for all integers $m, n > 1$.* This holds, for example, if t is an algebraic irrational number.

6. Remarks and questions

Polynomial solutions. A special case of polynomials P and Q for which there exist polynomial solutions to (2) (that is, the existence of *polynomials* f and g for which $f \circ P = g \circ Q$) is when P and Q commute under composition. J.F. Ritt [R2] showed, roughly, that P and Q commute only when they are both powers of z or when they are both Chebyshev polynomials of the first kind, or when they are both iterates (under composition) of the same polynomial (here, for simplicity of presentation, we have omitted some roots of unity as well as conjugacy involving linear polynomials).

In another paper written a couple of years earlier, Ritt [R1] provides a useful framework for studying the general question of polynomial solutions to (2) (see [J] and [Fa] for related work). Following the terminology of factorization of integers, Ritt called a polynomial *composite* if it can be represented as the composition of two polynomials each of degree larger than 1, and *prime* otherwise. His solution to the problem of representing a polynomial as a composition of prime polynomials is summarized in A and B below:

A. Any two decompositions of a given polynomial into prime polynomials contain the same number of polynomials; the degrees of the polynomials in one decomposition are the same as those in the other, except, perhaps, for the order in which they occur.

   Two decompositions of a polynomial F into the same number of polynomials,

$$F = \varphi_1 \varphi_2 \cdots \varphi_r, \qquad F = \psi_1 \psi_2 \cdots \psi_r$$

are called *equivalent* if there exist $r-1$ polynomials of the first degree

$$\lambda_1, \lambda_2, \ldots, \lambda_{r-1}$$

such that

$$\psi_1 = \varphi_1\lambda_1, \psi_2 = \lambda_1^{-1}\varphi_2\lambda_2, \ldots, \psi_r = \lambda_{r-1}^{-1}\varphi_r.$$

B. If F has two distinct decompositions into prime polynomials, we can pass from either to a decomposition equivalent to the other by repeated switchings of adjacent prime polynomials as follows: $\varphi_i\varphi_{i+1}$ may be replaced by $\psi_i\psi_{i+1}$ precisely when

$$\varphi_i = \lambda_1\pi_1\lambda_2, \qquad \varphi_{i+1} = \lambda_2^{-1}\pi_2\lambda_3$$

and

$$\psi_i = \lambda_1\xi_1\lambda_2, \qquad \psi_{i+1} = \lambda_2^{-1}\xi_2\lambda_3$$

and the pairs π_1, π_2 and ξ_1, ξ_2 are of the following three types:
(a) $\pi_1 = T_m$, $\pi_2 = T_n$, $\xi_1 = T_n$, $\xi_2 = T_m$, (Chebyshev polynomials),
(b) $\pi_1 = z^m$, $\pi_2 = z^r g(z^m)$, $\xi_1 = z^r[g(z)]^m$, $\xi_2 = z^m$, (g any polynomial),
(c) $\pi_1 = z^r[g(z)]^m$, $\pi_2 = z^m$, $\xi_1 = z^m$, $\xi_2 = z^r g(z^m)$, (g any polynomial).

Returning now to our question, suppose that P and Q are given and we seek polynomials f and g for which $f \circ P = g \circ Q$. But then, one must be able to go from the decomposition $f \circ P$ to $g \circ Q$ by means of the above "switchings" (a), (b), and (c). Thus, the question of the existence of polynomial solutions can be answered by studying the decompositions of P and Q.

By way of example, consider $P(x) := z^2(z^3+1)$ and $Q(z) := z^3$. P and Q do not commute under composition, but there do exist polynomial solutions $f(z) := z^3$, $g(z) := z^2(z+1)^3$ to equation (2), in keeping with (c) above.

Rational solutions. Using the asymptotic formula for the roots z of $P(z) = P(\zeta)$ (ζ large) given by Lemma 3.1 and the remark following, one can solve many cases of the af Hällström problem mentioned in Section 4 above. We restrict ourselves to one rather simple result, generalizations are straightforward and may be left to the reader.

Theorem 6.1. *Let $P(z) = z^p + az^{p-1} + \cdots$, $Q(z) = z^p + bz^{p-1} + \cdots$ be polynomials such that $p \geq 2$ and $a \neq b$. Then $\mathscr{M}_P \cap \mathscr{M}_Q$ contains no nonconstant rational function.*

Proof. By estimates from Section 3, for each $j \in \{1, 2, \ldots, p\}$, $P(z) = P(\zeta)$ has a root

$$z_j = \omega_j\zeta + \frac{P(\zeta) - P(\omega_j\zeta)}{P'(\omega_j\zeta)} + \mathscr{O}\big(|\zeta|^{-1}\big),$$

where $\omega_j = e^{2\pi ij/p}$. Simplifying, we get

$$z_j = \omega_j\zeta + (a/p)(\omega_j - 1) + \mathscr{O}\big(|\zeta|^{-1}\big). \tag{26}$$

In other words, ζ is P-equivalent to each of the numbers z_j given by (26), and in like manner it is Q-equivalent to each of the numbers

$$z'_j = \omega_j\zeta + (b/p)(\omega_j - 1) + \mathscr{O}\big(|\zeta|^{-1}\big).$$

Thus, by one P-step with $j = 1$ followed by a Q-step with $j = p-1$, ζ is transformed to (here, we denote ω_1 by ω, so $\omega_{p-1} = \overline{\omega}$)

$$\overline{\omega}\big(\omega\zeta + (a/p)(\omega - 1)\big) + (b/p)(\overline{\omega} - 1) + \mathscr{O}\big(|\zeta|^{-1}\big) = \zeta + \Big(\frac{a-b}{p}\Big)(1-\overline{\omega}) + \mathscr{O}\big(|\zeta|^{-1}\big).$$

Consequently, by a P-step and then a Q-step, ζ is shifted to $\zeta + c$, where c is the nonzero number $\big((a-b)/p\big)(1-\overline{\omega})$, modulo a $\mathscr{O}\big(|\zeta|^{-1}\big)$ error. It follows at once that, given any N, if ζ is chosen large enough, there are points

$$\zeta, \zeta + c + \mathscr{O}\big(|\zeta|^{-1}\big),\ \zeta + 2c + \mathscr{O}\big(|\zeta|^{-1}\big), \cdots, \zeta + Nc + \mathscr{O}\big(|\zeta|^{-1}\big)$$

which are all distinct and are $\equiv_{P,Q}$ to ζ. Clearly, no nonconstant rational function can take the same value at all these points, once N is chosen sufficiently large. □

In a similar way, we could show under the same hypotheses that $\mathscr{M}_P \cap \mathscr{M}_Q$ cannot contain a nonconstant meromorphic function of sufficiently small order.

Uniformization. The functional equation (2) can also be looked at from a slightly different point of view, very close to that in the papers of Marty, Shimizu and af Hällström already mentioned. Namely, the equations $w_1 = P(z)$, $w_2 = Q(z)$ can be viewed as parametric equations of an algebraic variety

$$V := \big\{(w_1, w_2) \in \mathbf{C}^2 : \varphi(w_1, w_2) = 0\big\}$$

of complex dimension one in $\mathbf{C}^2$. (Here φ is some polynomial in w_1, w_2. Because P and Q are polynomials, this variety (curve) is of genus zero, but it is more natural now to allow φ to be an arbitrary irreducible polynomial).

Our basic question then takes the form: *do there exist nonconstant meromorphic (or entire) functions f, g on $\mathbf{C}$ such that $f(w_1) = g(w_2)$ whenever $(w_1, w_2) \in V$?* Yet another way to look at matters is this: The equation $\varphi(w_1, w_2) = 0$ can be solved for $w_2 = A(w_1)$, where A is an (in general, multi-valued) algebraic function. That there exist nonconstant meromorphic functions f and g with $f(w_1) = g\big(A(w_2)\big)$ is the same as requiring: *to the given algebraic function A, there corresponds a nonconstant meromorphic function g on $\mathbf{C}$ such that $g \circ A$ remains single-valued under all analytic continuations.* The question then arises: Which algebraic functions admit such a "uniformizing" g? Notice that this is quite a different problem from the classical uniformization problem.

References

[AS] Aharonov, D., and H.S. Shapiro: Domains on which analytic functions satisfy quadratic identities. - J. Analyse Math. 30, 1976, 39–73.

[Fa] Fatou, P.: Sur les équations fonctionelles. - Bull. Soc. Math. France, XLVII–XLVIII, 1919–1920.

[Fl] Flatto, L.: A theorem on level curves of harmonic functions. - J. London Math. Soc. (2) 1, 1969, 470–472.

[FNS] Flatto, L., D.J. Newman, and H. S. Shapiro: The level curves of harmonic functions. - Trans. Amer. Math. Soc. 123, 1966, 425–436.

[FG] Fuchs, W.H.J., and F. Gross: Generalization of a theorem of A. and C. Rényi on periodic functions. - Acta Sci. Math. 32, 1971, 83–86.

[H1] af Hällström, G.: Über Substitutionen, die eine rationale Funktion invariant lassen. - Acta Acad. Åbo. Ser. B XV.6, 1946, 3–44.

[H2] af Hällström, G.: Über die Automorphiefunktionen meromorpher Funktionen. - Acta Acad. Åbo. Ser. B XVI.4, 1949, 3–27.

[J] Julia, G.: Mémoire sur la permutabilité des fractions rationelles. - Annales de l'Ecole Normale Supérieure, 39, 1922, 131–215.

[L] Langman, H.: Conformal transformations of period n and groups generated by them. - Amer. J. Math. XLIV, 1922, 54–86.

[M1] Marty, F.: Recherches sur la répartition des valeurs d'une fonction méromorphe. - Ann. Fac. Sci. Toulouse Math. (3) 23, 1931, 183–261.

[M2] Marty, F.: Sur le groupe d'automorphie de certaines fonctions entières. - C. R. Acad. Sci. Paris 195, 1932, 414–416.

[RR] Rényi, A., and C. Rényi: Some remarks on periodic entire functions. - J. Analyse Math. 14, 1965, 303–310.

[R1] Ritt, J.F.: Prime and composite polynomials. - Trans. Amer. Math. Soc. 23, 1922, 51–66.

[R2] Ritt, J.F.: Permutable rational functions. - Trans. Amer. Math. Soc. 25, 1923, 399–448.

[Sha] Shapiro, H.S.: The functional equation $f(P(z)) = g(Q(z))$ and a problem of A. and C. Rényi. - Studia Sci. Math. Hungar. 1, 1966, 255–259.

[Shi1] Shimizu, T.: On the fundamental domains and the groups for meromorphic functions I–II. - Japan. J. Math. VIII, 1931, 175–304.

[Shi2] Shimizu, T.: On the existence of meromorphic functions which are automorphic with respect to some function-groups. - Proc. Phys.-Math. Soc. Japan (3) 13:10, 1931, 292–296.

[Shi3] Shimizu, T.: On the linear functions of automorphism for meromorphic functions. - Tôhoku Math. J. 38, 1933, 219–224.

[Si] Siegel, C.L.: Iteration of analytic functions. - Ann. of Math. 43, 1942, 607–612 (reprinted in his Werke, Vol. II).

Received 9 April 1992