ON RITT'S FACTORIZATION OF POLYNOMIALS

A.F.BEARDON AND T.W.NG *

ABSTRACT

Ritt has shown that any complex polynomial p can be written as the composition of polynomials $p_1, \dots, p_m$, where each p_j is prime in the sense that it cannot be written as a non-trivial composition of polynomials. The factors p_j are not unique but the number m of them is, as is the set of the degrees of the p_j. Here we extend Ritt's theory and, in particular, we introduce a third invariant of the decomposition.

1. *Introduction*

The composition $p \circ q$ of two polynomials p and q is the map $z \mapsto p(q(z))$, and a non-linear polynomial p is said to be *prime* (with respect to composition) if $p = p_1 \circ p_2$ implies that p_1 or p_2 is a linear polynomial. If $p = p_1 \circ \cdots \circ p_m$, we say that the p_j are *factors* of p, and that $p_1 \circ \cdots \circ p_m$ is a *decomposition*, or a *factorization* of p. It is clear (by induction) that any polynomial p can be expressed in the form $p_1 \circ \cdots \circ p_m$, where each p_j is prime, and we say that this is a *prime decomposition* of p. The fundamental results in this area were first proved by Ritt ([6]) who showed that although a prime decomposition of a polynomial p is not unique, there are two invariants of a prime decomposition of a polynomial p, namely the number m of factors in the decomposition, and also the set of the degrees of these factors. Later in this paper we shall introduce a third invariant of the decomposition.

Ritt also described the extent of the non-uniqueness of a prime decomposition by showing how to pass from one prime decomposition of p to another. Although this result is fundamental and frequently quoted, it seems difficult to apply (and we are not aware of any applications of it in the literature). One of our aims here is to clarify and extend this result and, incidentally, to introduce some terminology and notation which we have found to be useful when discussing this topic.

Ritt showed that any prime factorization of a given polynomial p can be obtained from any other by a sequence of applications of any of the following three types (or their inverses): starting with the factorization $p_1 \circ \cdots \circ p_m$, we can (for any j)

(1) replace p_j and p_{j+1} by $p_j \circ \ell$ and $\ell^{-1} \circ p_{j+1}$, where ℓ is any linear polynomial, or

(2) replace p_j and p_{j+1} by p_{j+1} and p_j, where p_j and p_{j+1} are Tchebychev polynomials, or

(3) replace z^k and $z^r g(z^k)$ (which are p_j and p_{j+1}, respectively) by $z^r g(z)^k$ and z^k (which are q_j and q_{j+1}, respectively), where r and k are integers, and g is a polynomial.

We shall call these transformations (and their inverses) *Ritt transformations* of types 1, 2 and 3,

* Dr T.W.Ng thanks the Croucher Foundation for their support during the preparation of this paper.

1992 *Mathematics subject classification* 30D05

respectively. A few remarks may be helpful here. First, in each case we replace p_j and p_{j+1} by q_j and q_{j+1}, where $p_j \circ p_{j+1} = q_j \circ q_{j+1}$. Next, the transformation (1) guarantees that any polynomial p has *infinitely many* different prime decompositions, and our aim is eventually to show that in essence the infinity of possible decompositions can only arise from the use of (1). All pairs of Tchebychev polynomials $T_2, T_3 \ldots$. commute with each other ($T_a \circ T_b = T_b \circ T_a$) and this is the reason for (2). The transformation (3) expresses the fact that (with the obvious notation)

$$z^k \circ z^r g(z^k) = \left[z^r g(z^k)\right]^k = z^r g(z)^k \circ z^k, \tag{1.1}$$

and this includes the case $z^k \circ z^r = z^r \circ z^k$. We remark in passing that Ritt's theorems imply that $z^r g(z^k)$ is composite if and only if $z^r g(z)^k$ is composite (and this seems difficult to prove directly). Indeed, as (1.1) holds, the number of prime factors on each side of (1.1) is the same; thus the number of prime factors of $z^r g(z^k)$ is the same as that of $z^r g(z)^k$.

Now these rules are a little less transparent, and a little less independent, than may appear at first sight. First, we note that (3), which is stated in its conventional form, is rather loosely defined for the r and g are not uniquely determined by the form $z^r g(z^k)$; for instance, if $g(0) = 0$, we can equally well write this expression in the form $z^{r+k} h(z^k)$, where $h(z) = g(z)/z$. Next, $T_2(z) = 2z^2 - 1$ differs by a linear factor from z^2, so that it in some circumstances it is possible to apply (2) to T_2, then (1), and then (on what is essentially the same factor) (3). These observations perhaps show why it is difficult to use Ritt's result. We remark that T_2 is the only even Tchebychev polynomial to occur in a prime decomposition (because for $n \geq 2$, $T_{2n}(z)$ is an even function and hence composite). On the other hand, as each Tchebychev polynomial T_{2n+1} is an odd function it is of the form $zg(z^2)$ and so it may be possible to apply (3). However, T_{2n+1} is only of the form $z^r g(z^k)$ when $r = 1$ and $k = 2$, and it is not of the form $z^r g(z)^k$ unless $r = k = 1$ (for T_{2n+1} has only real, and simple, zeros).

It seems to be the case that the Ritt transformation (1) plays a different role to the transformations (2) and (3), although we know of nothing in the literature to suggest that this might be so. In general, we expect (2) and (3) to be applicable in only rather exceptional cases, whereas (1) is always applicable, in infinitely many ways, and regardless of the particular form of the p_j. It is the rule (1), for example, that forces each p to have *infinitely many* distinct prime factorizations, whereas there is a sense in which the transformations (2) and (3) can only contribute to a *finite number* of possible distinct factorizations; this does not appear to have been noticed (or at least discussed) before. It might be advantageous then to somehow ignore (1), and focus on the impact of (2) and (3) only. We shall show how to do this, and we obtain the following result.

THEOREM 1.1. *Given a polynomial p, there are a finite number of factorizations of p, which we denote symbolically by $\mathcal{P}_1, \ldots, \mathcal{P}_m$, such that any factorization of p is of the form*

$$(p_1 \circ \ell_1) \circ (\ell_1^{-1} \circ p_2 \circ \ell_2) \circ \cdots \circ (\ell_{n-1} \circ p_n),$$

where $p_1 \circ \cdots \circ p_n$ is one of the factorizations $\mathcal{P}_j$ of p, and the ℓ_j are linear polynomials.

This result complements Ritt's Theorem. It is stronger because Ritt's result does not seem to imply directly that up to the insertion of linear factors, there are only *finitely many* possible factorizations, and in any case our result applies to all factorizations (whether prime or not). It is weaker as it does not explicitly link any two possible (prime) factorizations of p. We shall prove Theorem 1.1 in Section 2, where we shall also present a more formal language which seems to be useful when giving a careful discussion of these ideas.

In the remaining sections of this paper we shall obtain a significant extension of the results in [3] by applying our ideas to polynomials that are invariant under a rotation and, as we shall see shortly, these ideas lead to a third invariant of prime decompositions. The paper [3] starts with the question 'if f and g are entire functions whose composition $f \circ g$ is even, must f or g be even?' and contains the following two results on polynomials.

THEOREM A. *Suppose that p and q are polynomials with $q(0) = 0$, and $p \circ q$ even. Then either* (i) *q is even, or* (ii) *p is even and q is odd.*

THEOREM B. *If some iterate of a polynomial p is even, then p is even.*

Now an even function is simply a function that is invariant under the Euclidean rotation $z \mapsto -z$, and accordingly, one should seek (and perhaps expect) analogous results for a general finite group of rotations. These results should be of the form that if the composition $p_1 \circ \cdots \circ p_m$ is invariant under some rotation group, then some (or all) of the composition factors p_j should also exhibit some rotational invariance, and that collectively, they should exhibit enough invariance to account for that of p. Of course, if $p \circ q$ and q are even then they are invariant under the *same* rotation group (containing just the identity and $z \mapsto -z$); we shall see that this restriction is uneccesary (it is forced on one by over-emphazing the concept of 'even' functions). The analysis of these issues clarifies the role of the assumption $q(0) = 0$ which is emphasized several times in [2] and [3]. To some extent the discussion in [2] (pp.231-233) generalises that in [3], but it is still confined to rotations about the origin. Here we give a simpler, but more general, discussion based on the ideas on Ritt's Theorems.

By a *rotation group* we mean a finite group of Euclidean rotations in $\mathbb{C}$. If α and β are rotations, then the commutator $\alpha\beta\alpha^{-1}\beta^{-1}$ is a translation unless α and β have a common fixed point, and it follows from this that elements in a rotation group Γ have a common fixed point; thus a rotation group is simply a finite cyclic group of Euclidean rotations about some point. Note that given any nonconstant polynomial p, the set $\Gamma(p)$ of Euclidean isometries γ such that $p \circ \gamma = p$ is a group. As p cannot be periodic, $\Gamma(p)$ is a rotation group and we deduce that if γ_1 and γ_2 are nontrivial rotations that leave p invariant, then they have the same fixed point. We shall use the notation $\Gamma(p)$ throughout this paper; likewise, we shall use $\zeta(p)$ to denote the common fixed point of elements of $\Gamma(p)$ (when this group is nontrivial).

Now consider the polynomial

$$p(z) = z^2 \circ z^3 \circ z^5 \circ z^7 \circ (z+1) \circ z^2 = (z^2+1)^{210} = 1 + 210z^2 + \cdots.$$

As p is even it is invariant under $z \mapsto -z$, so that if p is also invariant under a rotation γ, then $\gamma(0) = 0$ and hence $\gamma(z) = \lambda z$, say. Clearly $\lambda^2 = 1$ so we deduce that $\Gamma(p) = \{I, \sigma\}$, where (here and elsewhere) I is the identity map and $\sigma(z) = -z$. On the other hand, the composition factors of p have rotational symmetries of orders 2, 3, 5, 7, 1 and 2 so that in this case the collective symmetries of the composition factors of p far exceeds the symmetry of the composite polynomial p. Roughly speaking, this example shows that some of the symmetries of the polynomials p_j can be destroyed in passing to the composition $p_1 \circ \cdots \circ p_m$. Our next result shows (roughly speaking) that we cannot create additional symmetries by passing to a composition; for example, if none of the composition factors p_j have any rotational symmetry, then neither does their composition $p_1 \circ \cdots \circ p_m$.

THEOREM 1.2. *Suppose that $p_1, \ldots, p_m$ are nonconstant polynomials, and $\Gamma(p_1), \ldots, \Gamma(p_m)$, and $\Gamma(p_1 \circ \cdots \circ p_m)$ have orders $k_1, \ldots, k_m$ and k, respectively. Then k divides $k_1 \cdots k_m$.*

In Section 3 we shall give an example to show that the centres of rotation $\zeta(p_1), \ldots, \zeta(p_m)$ in Theorem 1.2 can be distinct.

We are now in a position to describe the third invariant of a prime decomposition of a polynomial. Suppose that

$$p_1 \circ \cdots \circ p_m = p = q_1 \circ \cdots \circ q_m \tag{1.2}$$

are two different prime decmpositions of the polynomial p. Then associated with these we have the rotation groups $\Gamma(p_1), \ldots, \Gamma(p_m)$, and their orders, say $k_1, \ldots, k_m$, respectively, and similar groups for the q_j, say with orders $k_1', \ldots, k_m'$, respectively. The third invariant is the vector $(k_1, \ldots, k_m)$ modulo the natural action of permutations.

THEOREM 1.3. *If the two decompositions in* (1.2) *are prime decompositions of p and if the k_i and k_j' are as defined above, then the vector $(k_1', \ldots, k_m')$ is obtained from $(k_1, \ldots, k_m)$ by a permutation of its components.*

Finally, we shall prove the following self-explanatory 'structural' result on the decomposition of any polynomial.

THEOREM 1.4. *Suppose that p is a nonconstant polynomial and that $\Gamma(p)$ has order k. Then we can write $p = \ell_1 \circ p_1 \circ \ell_2 \circ \cdots \circ \ell_m \circ p_m \circ \ell_{m+1}$, where each ℓ_i is linear, and for each j, either $p_j(z) = z^{k_j}$, where k_j is prime, or p_j is a prime polynomial that is not invariant under any nontrivial rotation group (in which case we put $k_j = 1$). Again, k divides $k_1 \cdots k_m$.*

We prove Theorem 1.1 in Section 2. The remaining results involve rotation groups, and Section 3 is devoted to some general remarks about polynomials invariant under a rotation. The proofs of Theorems 1.2, 1.3 and 1.4 are given in Sections 4, 5 and 6, respectively. In Section 7 we show that our methods can be successfully applied to such problems as those discussed in [2] and [3]. Our intention here is to illustrate our methods by discussing Theorems A and B rather than to

give a complete discussion. We end, in Section 8, with a brief remark about factorization of entire functions.

2. *The proof of Theorem 1.1*

We say that a polynomial p is *normalised* if it is monic with $p(0) = 0$. Note that if $p = q \circ r$, and if p and r are normalised, then so too is q. We begin with the following two lemmas.

LEMMA 2.1. *Suppose that p, q and r are normalised polynomials such $p = q \circ r$. Then r is uniquely determined by p and $\deg(r)$.*

LEMMA 2.2. *Suppose that p_1, p_2, q_1 and q_2 are polynomials such that $p_1 \circ q_1 = p_2 \circ q_2$, and $\deg(q_1) = \deg(q_2)$. Then there is a linear polynomial α such that $q_2 = \alpha \circ q_1$.*

Lemma 2.1 is discussed in [4] (in a rather more complicated form, and as a computer algorithm), and stated in [7] (p.285), although it is not apparent from either of these that the result follows simply by equating coefficients in an identity. We give a proof by this method below. Lemma 2.2 is stated in [5] (p.221) in the case when $p_1 = p_2$ (where this simpler case of the much harder problem for rational functions is discussed).

The proof of Lemma 2.1 Suppose that p, q and r have given degrees n, k and m, respectively, so that $n = km$. As $p = q \circ r$, and q is normalised,

$$p(z) = r(z)^k + O\big(r(z)^{k-1}\big) = r(z)^k + O(z^{n-m}),$$

near ∞, and we have to show that r is uniquely determined by p. We write

$$p(z) = a_1 z + a_2 z^2 + \cdots + a_{n-1} z^{n-1} + a_n z^n, \quad (a_n = 1),$$
$$r(z) = c_1 z + c_2 z^2 + \cdots + c_{m-1} z^{m-1} + c_m z^m, \quad (c_m = 1),$$

and the proof is simply by comparing coefficients. As $p(z) = [r(z)]^k + O(z^{km-m})$, the Multinomial Theorem gives

$$p(z) = \sum_{u_1,\ldots,u_m} \frac{k!}{u_1! \cdots u_m!} (c_1 z)^{u_1} \cdots (c_m z^m)^{u_m} \; + \; O(z^{km-m}),$$

where the sum is over all $(u_1, \ldots, u_m)$ with non-negative integers u_i that satisfy $u_1 + \cdots + u_m = k$. Equating coefficients of $z^{km-m+1}, \ldots, z^{km-m+(m-1)}$, we obtain

$$a_{km-m+\ell} = \sum_{u_1,\ldots,u_m} \frac{k!}{u_1! \cdots u_m!} c_1^{u_1} \cdots c_m^{u_m}, \quad \ell = 1, \ldots, m-1,$$

where this sum is over all $(u_1, \ldots, u_m)$ with $u_i \geq 0$, $u_i \in \mathbb{Z}$, $\sum_j u_j = k$ and $\sum_j j u_j = km - m + \ell$ or, equivalently, with $u_i \geq 0$, $u_i \in \mathbb{Z}$, and

$$\begin{aligned} m - \ell &= km - \sum_j j u_j \\ &= m \sum_j u_j - \sum_j j u_j \\ &= (m-1)u_1 + (m-2)u_2 + \cdots + 1 \cdot u_{m-1}, \end{aligned} \tag{2.1}$$

and

$$u_m = k - (u_1 + \cdots + u_{m-1}). \tag{2.2}$$

It is important to note that the set of $(u_1, \ldots, u_m)$ over which we are summing here depends on ℓ, so we shall denote this set by $\mathcal{U}(\ell)$.

Now fix a value of ℓ, and consider a general $(u_1, \ldots, u_m)$ in $\mathcal{U}(\ell)$. Equation (2.1) implies that $m-\ell \geq (m-j)u_j$, so that when $j < \ell$ we have $u_j < 1$ and hence $u_j = 0$; thus $u_1 = \cdots = u_{\ell-1} = 0$. Similarly, u_ℓ is 0 or 1, and when $u_\ell = 1$ then $u_{\ell+1} = \cdots = u_{m-1} = 0$. Using (2.2) we deduce that $\mathcal{U}(\ell)$ contains the vector $(0, \ldots, 0, 1, 0, \ldots, 0, k-1)$ with 1 in the ℓ-th place, and that all other vectors in $\mathcal{U}(\ell)$ have $u_1 = \cdots = u_\ell = 0$. This shows that

$$a_{k\,m-m+\ell} = kc_\ell c_m^{k-1} + \sum_{u_{\ell+1}, \ldots, u_m} \frac{k!}{u_{\ell+1}! \cdots u_m!}\, c_{\ell+1}^{u_{\ell+1}} \cdots c_m^{u_m}, \tag{2.3}$$

where $c_m = 1$, and where this sum is taken over all vectors $(u_{\ell+1}, \ldots, u_m)$ with $u_i \geq 0$, $u_i \in \mathbb{Z}$, and

$$u_{\ell+1} + \cdots + u_m = k,$$
$$(\ell+1)u_{\ell+1} + \cdots + mu_m = km - m + \ell.$$

When $\ell = m-1$ the sum in (2.3) is empty (for then, $u_m = k$ and $mk = km - 1$) and so is 0. Notice that c_ℓ occurs with a nonzero coefficient, and only in the first term on the right in (2.3); thus we can write (2.3) in the form

$$a_{k\,m-m+\ell} = kc_\ell + \Phi_{k,m,\ell}(c_{\ell+1}, \ldots, c_{m-1}, c_m),$$

where $\Phi_{k,m,\ell}$ is a polynomial whose coefficients depend only on k, m and ℓ. Now $c_m = 1$, and it follows that $c_{m-1}, c_{m-2}, \ldots c_1$ can be computed inductively (in this order) so that as promised, r is uniquely determined by p and m. This completes the proof of Lemma 2.1.

The proof of Lemma 2.2 Suppose that $p_1 \circ q_1 = p_2 \circ q_2 = f$, say, where $\deg(q_1) = \deg(q_2)$. There are (unique) linear polynomials α, β_1 and β_2 such that $\alpha \circ f$, $\beta_1 \circ q_1$ and $\beta_2 \circ q_2$ are all normalised. As

$$\alpha \circ f = \big(\alpha \circ p_j \circ \beta_j^{-1}\big) \circ \big(\beta_j \circ q_j\big),$$

for $j = 1, 2$, and as all three polynomials here are normalised, Lemma 2.1 implies that $\beta_1 \circ q_1 = \beta_2 \circ q_2$ and Lemma 2.2 follows.

We now introduce some formal terminology in preparation for our proof of Theorem 1.1. The ideas that follow were motivated by Ritt's theory but we stress that here we do not restrict ourselves to prime polynomials (or prime factors). Fix a positive integer m and let $\mathcal{S}_m$ be the set of all sequences $(p_1, \ldots, p_m)$, where each p_j is a polynomial. In general, we shall use the vector notation $P = (p_1, \ldots, p_m)$, $Q = (q_1, \ldots, q_m)$ and so on. Now define the relation $\sim$ on $\mathcal{S}_m$ by $(p_1, \ldots, p_m) \sim (q_1, \ldots, q_m)$ if and only if there exists j in $\{1, 2, \ldots, m-1\}$ such that

(i) $p_i = q_i$ if $i \neq j, j+1$, and

(ii) q_j and q_{j+1} are obtained from p_j and p_{j+1} by one of the Ritt transformations (1), (2) or (3), or their inverses.

The relation $\sim$ is reflexive and symmetric and so it induces an equivalence relation $\approx$ on $\mathcal{S}_m$ in the usual way by the existence of some P_i such that

$$P \approx Q \Longleftrightarrow P = P_1 \sim P_2 \sim \cdots \sim P_{k-1} \sim P_k = Q.$$

We say that P and Q are *Ritt equivalent* if $P \approx Q$, and we call a $\approx$-equivalence class a *Ritt class.* Note that if $(p_1, \ldots, p_m) \approx (q_1, \ldots, q_m)$, then $p_1 \circ \cdots \circ p_m = q_1 \circ \ldots \circ q_m$.

We remark that if $\mathcal{S}_m$ were to be taken as the set of all sequences $(p_1, \ldots, p_m)$, where each p_j is a *prime* polynomial, then the Ritt class containing $(p_1, \ldots, p_m)$ would be the collection of all prime decompositions of the polynomial $p_1 \circ \cdots \circ p_m$ (this is one of Ritt's Theorems), but again we stress that we are considering composite as well as prime polynomials here.

In order to distinguish the Ritt transformation (1) from (2) and (3) we define a second equivalence relation $\asymp$ on $\mathcal{S}_m$. First, we define a relation $\smile$ on $\mathcal{S}_m$ by saying that $(p_1, \ldots, p_m) \smile (q_1, \ldots, q_m)$ if and only if these two vectors are obtained from each other by a Ritt transformation of the type (1). This relation is symmetric and reflexive so it extends to an equivalence relation $\asymp$ on $\mathcal{S}_m$ exactly as above. Observe that P and Q, say, lie in the same $\asymp$-equivalence class if and only if we can convert P into Q by a finite sequence of applications of the rule (1); similarly, P and Q, say, lie in the same $\approx$-equivalence class if and only if we can convert P into Q by a finite sequence of applications of any of the rules (1), (2) and (3).

As $P \asymp Q$ implies that $P \approx Q$, the $\asymp$-equivalence classes partition each $\approx$-equivalence class, and so Theorem 1.1 can be restated in this notation as follows.

THEOREM 1.1A. *Each $\approx$-equivalence class is a finite union of $\asymp$-equivalence classes.*

Proof. Suppose first that

$$p_1 \circ \cdots \circ p_m \approx q_1 \circ \cdots \circ q_m, \tag{2.4}$$

where for each i,

$$\deg(p_i) = \deg(q_i). \tag{2.5}$$

Then $p_1 \circ (p_2 \circ \cdots \circ p_m) = q_1 \circ (q_2 \circ \cdots \circ q_m)$, so that by Lemma 2.2, there is a linear polynomial ℓ_1 such that

$$p_2 \circ \cdots \circ p_m = \ell_1 \circ q_2 \circ \cdots \circ q_m. \tag{2.6}$$

It follows that

$$q_1 \circ (q_2 \circ \cdots \circ q_m) = p_1 \circ p_2 \circ \cdots \circ p_m = p_1 \circ \ell_1 \circ (q_2 \circ \cdots \circ q_m)$$

from which we can deduce that $q_1 = p_1 \circ \ell$, and hence $p_1 = q_1 \circ \ell_1^{-1}$. From (2.6) we have

$$p_2 \circ \cdots \circ p_m = (\ell_1 \circ q_2) \circ \cdots \circ q_m,$$

and a repetition of the same argument gives $p_2 = \ell_1 \circ q_2 \circ \ell_2^{-1}$ and so on. This shows that if (2.4) and (2.5) hold, then $(p_1, \ldots, p_m) \asymp (q_1, \ldots, q_m)$.

It now follows that each $\approx$-equivalence class is the union of $\asymp$-equivalence classes, and that if $(p_1, \ldots p_m)$ and $(q_1, \ldots, q_m)$ lie in the same $\approx$-equivalence class but in different $\asymp$-equivalence classes, then (2.5) fails so that

$$\big(\deg(p_1), \ldots, \deg(p_m)\big) \neq \big(\deg(q_1), \ldots, \deg(q_m)\big). \tag{2.7}$$

Theorem 1.1 now follows easily because if p has a factorization $p_1 \circ \cdots \circ p_m$, say, then the degrees of the p_j are factors of $\deg(p)$ and so there are only a finite number of possibilities for the vector on the left hand side of (2.7).

3. *Polynomials invariant under rotations*

Suppose that a nonconstant polynomial p is invariant under a nontrivial rotation group Γ with fixed point z_0. If $\gamma \in \Gamma$, then $p'(z) = (p \circ \gamma)'(z) = p'\big(\gamma(z)\big)\gamma'(z)$, and as $\gamma'(z) \neq 0$ this implies that Γ acts as a permutation group on the set $C(p)$ of critical points of p. As $\gamma(z) = az + b$ for some a and b, γ fixes the arithmetic mean of $C(p)$, so this point must be z_0. It follows that $(n-1)z_0$ is the sum of the zeros of $p'(z)$, so that if

$$p(z) = a_0 + \cdots + a_n z^n, \tag{3.1}$$

where $a_n \neq 0$, then $z_0 = -a_{n-1}/na_n$. In addition, as p is invariant under Γ, p is not injective near z_0, thus z_0 is one of the critical points of p. These observations yield the following result.

LEMMA 3.1. *Suppose that p is given by* (3.1). *If p is invariant under some rotation group Γ with fixed point z_0, then $z_0 = -a_{n-1}/na_n$, and $p'(z_0) = 0$. Further, z_0 is the arithmetic mean of the critical points of p.*

Lemma 3.1 shows (for example) that if $a_1 \neq 0$ and $a_{n-1} = 0$, then $z_0 = 0$ and $p'(0) \neq 0$ so that p is not invariant under *any* rotation group.

Next, suppose that p is a nonconstant prime polynomial, and that $\Gamma(p)$ is generated by γ, where

$$\gamma(z) = z_0 + \omega(z - z_0), \quad \omega = \exp(2\pi i/k), \quad k \geq 2.$$

Now write $\tau(z) = z + z_0$ and $\rho(z) = \omega z$. Then $\gamma = \tau \circ \rho \circ \tau^{-1}$, so that $p \circ \tau$ is invariant under ρ. It is easy to see that this implies that $p \circ \tau(z) = q(z^k)$ for some polynomial q, and as $k \geq 2$ and p is prime, q must be of degree one and k must be prime. For each positive integer k we define $\sigma_k(z) = z^k$. Then $p \circ \tau = q \circ \sigma_k$, where τ and q are linear polynomials, and we have proved the next lemma.

LEMMA 3.2. *Suppose that p is a prime polynomial, and that $\Gamma(p)$ has order k, where $k \geq 2$. Then k is a prime integer, and there are linear polynomials α and β such that $p = \alpha \circ \sigma_k \circ \beta$. Further, we can take β to be a translation with $\beta^{-1}(0)$ the fixed point of $\Gamma(p)$.*

Finally, we shall need the following simple result.

LEMMA 3.3. *For any polynomial p, and any linear polynomial ℓ, $\Gamma(p \circ \ell) = \ell^{-1} \circ \Gamma(p) \circ \ell$ and $\Gamma(\ell \circ p) = \Gamma(p)$.*

Proof. The proof is easy; for example, $\gamma \in \Gamma(p \circ \ell)$ if and only if $p \circ \ell \circ \gamma = p \circ \ell$, and this is so if and only if $\ell \circ \gamma \circ \ell^{-1} \in \Gamma(p)$. The proof that $\Gamma(\ell \circ p) = \Gamma(p)$ is even easier and is omitted.

4. *The proof of Theorem 1.2*

In order to prove that k divides $k_1 \cdots k_m$ in Theorem 1.2 it clearly suffices to take $m = 2$ (the general case then following by induction). Thus it is sufficient to consider two polynomials p and q and show that if $\Gamma(p)$, $\Gamma(q)$ and $\Gamma(p \circ q)$ have orders u, v and k, respectively, then $k|uv$.

Suppose that γ generates $\Gamma(p \circ q)$. As $p \circ q = (p \circ q) \circ \gamma = p \circ (q \circ \gamma)$, Lemma 2.2 shows that there is a linear polynomial η such that $q \circ \gamma = \eta \circ q$. We deduce that $p \circ q = p \circ q \circ \gamma = p \circ \eta \circ q$, from which we deduce that $p = p \circ \eta$. As p is nonconstant, η must generate a discrete group, and as p cannot be periodic, this group must be a rotation group, say Σ_1 of order u_1. As Σ_1 is a subgroup of $\Gamma(p)$, we see that $u_1|u$. Next, as $q \circ \gamma = \eta \circ q$ we see that $q \circ \gamma^\ell = \eta^\ell \circ q$ for every integer ℓ. Taking $\ell = k$ (the order of $\Gamma(p \circ q)$), we see that $\eta^k = I$ (the identity) so that the order u_1 of η divides k. Finally, taking $\ell = u_1$, we see that q is invariant under the group generated by γ^{u_1}, and this has order $v_1 = k/u_1$. As $u_1|u$, $v_1|v$ and $u_1 v_1 = k$ we see that $k|uv$.

We remarked in Section 1 that we would give an example in which the points $\zeta(p_j)$ in Theorem 1.2 are distinct. This is so in the first of the two following examples.

Example 4.1 Let

$$p(z) = a + (z-a)^2, \quad q(z) = a + (z-b)^3, \quad r(z) = b + (z-c)^5,$$
$$\alpha(z) = a + \omega^{15}(z-a), \quad \beta(z) = b + \omega^5(z-b), \quad \gamma(z) = c + \omega(z-c),$$

where a, b and c are distinct, and $\omega = \exp(2\pi i/30)$. It is easy to verify that

$$p \circ \alpha = p, \quad q \circ \beta = \alpha \circ q, \quad r \circ \gamma = \beta \circ r.$$

First, this shows that p is invariant under the rotation α of order 2 about a. Next, as $q \circ \beta^2 = \alpha^2 \circ q = q$, we see that q is invariant under the rotation β^2 of order 3 about b. Similarly, as $r \circ \gamma^6 = \beta^6 \circ r = r$, r is invariant under the rotation γ^6 of order 5 about c. Finally, as $p \circ q \circ r \circ \gamma = p \circ q \circ \beta \circ r = p \circ \alpha \circ q \circ r = p \circ q \circ r$, we see that $p \circ q \circ r$ is invariant under the rotation γ of order 30 about c.

Example 4.2 Let $p(z) = 1 + z + z^3$ and $q(z) = z^2$. Here, $\Gamma(q) = \{I, \rho\}$, where I is the identity and $\rho(z) = -z$, and $\Gamma(p) = \{I\}$ (this follows from Lemma 4.1). As $p \circ q$ is even, $\rho \in \Gamma(p \circ q)$, and Theorem 1.2 shows that $\Gamma(p \circ q) = \{I, \rho\}$. This example shows that we may have $k_j = 1$ for some j in Theorem 1.2.

5. *The proof of Theorem 1.3*

Clearly we have only to establish the invariance described in Theorem 1.3 for each of the Ritt transformations of types 1,2 and 3. In each case we replace a pair of polynomials p_j, p_{j+1} by a pair q_j, q_{j+1}, whose rotation groups have orders $k_j, k_{j+1}, k'_j, k'_{j+1}$, respectively, and we have to show that the unordered pair (k_j, k_{j+1}) is the same as the unordered pair (k'_j, k'_{j+1}). We shall discuss each of the Ritt transformations in turn (although we shall change to a more convenient notation).

A Ritt transformation of Type 1 replaces a pair of polynomials p and q by the pair $p \circ \ell$ and $\ell^{-1} \circ q$, and the desired result in this case follows immediately from Lemma 3.3.

In the case of a Ritt transformation of Type 2 we replace two Tchebychev polynomials, say T_a and T_b, by T_b and T_a (in this order). In this case, $k_j = k'_{j+1}$ and $k_{j+1} = k'_j$ so there is nothing more to prove.

Finally, in a Ritt transformation of Type 3 we replace the polynomials z^k and $z^r g(z^k)$ by the (ordered) pair $z^r g(z)^k$ and z^k, and it suffices to show that the two finite groups

$$\Gamma\big(z^r g(z^k)\big), \quad \Gamma\big(z^r g(z)^k\big) \tag{5.1}$$

are of the same order. Notice that as this case only arises when a factor z^k is present, and as z^k is prime only if k is prime, we may assume here that k is a prime and $k \geq 2$. In addition, $r \geq 1$ (else $z^r g(z^k)$ is not a prime factor) and $\gcd(r, k) = 1$ (again, else $z^r g(z^k)$ is not a prime factor). Thus to complete our proof of Theorem 1.3 we have only to show that the two groups in (5.1) have the same order under the conditions $r \geq 1$, $k \geq 2$ and is prime, and $\gcd(r, k) = 1$.

Suppose now that $z^r g(z^k)$ is invariant under a nontrivial rotation, say $\gamma(z) = z_0 + \omega(z - z_0)$, where ω $(\neq 1)$ is a root of unity. Then if we write

$$z^r g(z^k) = a_0 + a_1 z + \cdots + a_{n-1} z^{n-1} + a_n z^n,$$

we have $a_{n-1} = 0$ (as $k \geq 2$) so that, from Lemma 3.1, $z_0 = 0$ and $\gamma(z) = \omega z$. As (in this discussion) $z^r g(z^k)$ is prime, Lemma 3.2 shows that there is a linear polynomial α and an integer s such that $z^r g(z^k) = \alpha(z^s)$, and as $r \geq 1$ we have $\alpha(0) = 0$ so that for some constant μ, $z^r g(z^k) = \mu z^s$. This shows that $g(z) = \lambda z^t$, say, and using this we easily find that

$$z^r g(z)^k = \lambda^{k-1}[z^r g(z^k)].$$

Lemma 3.3 now shows that the two groups in (5.1) are identical. To summarise, we have shown that if $\Gamma\big(z^r g(z^k)\big)$ is nontrivial, then the two groups in (5.1) are identical.

We shall now show that if $\Gamma\left(z^r g(z)^k\right)$ is nontrivial, then again the two groups in (5.1) are identical. It then follows that *in all cases* the two groups in (5.1) are identical and this will complete our proof. We suppose, then that $z^r g(z)^k$ is invariant under the nontrivial rotation $\gamma(z) = z_0 + \omega(z - z_0)$. If $r \geq 2$ then Lemma 3.1 again shows that $z_0 = 0$ and the argument goes through exactly as before. The remaining possibility, namely that $r = 1$, seems to need a special argument. We suppose now that $r = 1$ and that $zg(z)^k$ is invariant under γ; then

$$\frac{z}{\gamma(z)} = \left[\frac{g(\gamma(z))}{g(z)}\right]^k.$$

The left hand side here is either constant or a Möbius map, and as $k \geq 2$, it must be constant. Thus $\gamma(z) = \omega z$ and the proof is again as before.

6. *The proof of Theorem 1.4*

We take any polynomial p and write it as a composition $p_1 \circ \cdots \circ p_m$, where each p_j is prime. If any p_j is invariant under some nontrivial rotation group then, by Lemma 3.2, we can replace it by a composition of the form $\ell_1 \circ q \circ \ell_2$, where the ℓ_i are linear, and this leads to the form described in Theorem 1.4. The fact that $k | k_1 \cdots k_m$ follows directly from Theorem 1.2.

7. *The results in [2] and [3]*

We begin by using the ideas above to prove Theorem B, namely that if some iterate of a polynomial p is even then p is even. First, Lemma 2.2 implies that if

$$p_1 \circ \cdots \circ p_m = q_1 \circ \cdots \circ q_m,$$

where all of the polynomials $p_1, \ldots, p_m, q_1, \ldots, q_m$ are of the same degree, then there must exist linear polynomials ℓ_i such that

$$q_1 = p_1 \circ \ell_1, \quad q_2 = \ell_1^{-1} \circ p_2 \circ \ell_2, \quad \ldots, \quad q_m = \ell_{m-1} \circ p_m. \tag{7.1}$$

Suppose now that the m-th iterate p^m is even; then $p \circ \cdots \circ p = p \circ \cdots \circ p \circ (p \circ \sigma)$, where $\sigma(z) = -z$, and where there are m factors on each side. According to (7.1), we can now write

$$p = p \circ \ell_1, \quad p = \ell_1^{-1} \circ p \circ \ell_2, \quad \ldots, \quad (p \circ \sigma) = \ell_{m-1} \circ p,$$

where each ℓ_i is linear.

Now $\ell_1 \in \Gamma(p)$, and $\Gamma(p)$ is contained in $\Gamma(p^m)$; thus $\ell_1 \in \Gamma(p^m)$. However, $\sigma \in \Gamma(p^m)$ and this forces ℓ_1 to be a rotation about the origin. We write $\ell_1(z) = \omega z$, where $|\omega| = 1$. Now suppose that $\Gamma(p)$ has order k and $\Gamma(p^m)$ has order s. Then $k | s$ and (from Theorem 1.2 with $p_j = p$ for all j) $s | k^m$. As $\sigma \in \Gamma(p^m)$ we see that s is even; thus k is even, so p is an even function and this is Theorem B.

It is an immediate consequence of Theorem 1.2 that if p^n is invariant under some nontrivial rotation group Γ_n of order v, say, then p is invariant under some nontrivial rotation group Γ_1 of order u, say, where $v|u^n$. As $\Gamma_1 \subset \Gamma_n$, these groups must have the same centres of rotation, and this shows why it is not necessary to assume that $p(0) = 0$ in the proof of Theorem B, [3] (see the remark on p.417). If v_1 is the product of the distinct prime factors of v, then $v_1|u^n$ so that $v_1|u$; thus $\Gamma(p)$ has order at least v_1. This contains Theorem B (which is the case $v_1 = 2$) and much more.

Consider now two nonconstant polynomials p and q, and suppose that $\Gamma(p)$, $\Gamma(q)$ and $\Gamma(p \circ q)$ have orders u, v and k, respectively, where k is prime. As $k|uv$ either $k|u$ or $k|v$. If $k|u$ then p has at least the same order of symmetry as $p \circ q$. Suppose now that $k|v$ and that γ generates $\Gamma(p \circ q)$ (so that γ has order k). Then (as before) there is a linear polynomial η such that $\eta \circ q = q \circ \gamma$, and p is invariant under the rotation group generated by η. The compatibility condition $q \circ \gamma = \eta \circ q$ implies that $\eta^k \circ q = q \circ \gamma^k = q$ so that η has order t, say, where $t|k$. Writing $k = st$, we note that $q = \eta^t \circ q = q \circ \gamma^t$ so that q is invariant under the rotation γ^t of order s. To summarise, p is invariant under η of order t, q is invariant under γ^t of order s, where $st = k$, and we have the compatibility condition $\eta \circ q = q \circ \gamma$ (which implies that $p \circ q$ is invariant under γ). All of this is *without* any assumption corresponding to the assumption $q(0) = 0$ made in [3].

The compatibility condition $q \circ \gamma = \eta \circ q$ implies (in general) that if γ fixes z^* then η fixes $q(z^*)$, so that the assumption $q(0) = 0$ (in [3]) is simply the requirement that η and γ have a common fixed point. If we too make this assumption, then $\eta = \gamma^r$, say, so that $rt = k$, whence $r = s$. In this case we see that p is invariant under γ^s of order t, q is invariant under γ^t of order s, where $st = k$, and $\gamma^s \circ q = q \circ \gamma$. Theorem A is the case $k = t = 2$ and $s = 1$ and $\gamma(z) = -z$; then $\eta = \gamma$ so that p is even, and the compatibility condition is then $\gamma \circ q = q \circ \gamma$ which says that q is odd.

Finally, we comment on Section 4 (pp.231-233) in [2]. Here, ω is a primitive N-th root of unity, where N is prime, and, writing $\gamma(z) = \omega z$, a function f is said to be *cyclic* if (in our simpler notation) for some k, $\gamma^k \circ f = f \circ \gamma$. Suppose now that p and q are nonconstant polynomials with $\gamma^k \circ p \circ q = p \circ q \circ \gamma$. Then, as before, there is a linear polynomial η such that $p \circ \eta = p$ and $\eta \circ q = q \circ \gamma$. As before, the order of η divides N, and as N is prime we have either $\eta = I$ or η has order N. With $q(0) = 0$ (an assumption made in [2]) we see that $\eta = I$ or η and γ generate the same group. Thus q is γ-invariant, or $\eta = \gamma^t$, say in which case $\gamma^t \circ q = q \circ \gamma$ so that q is cyclic. This is stronger than Proposition 4.1 in [2].

8. *A remark*

There is a substantial theory of factorization of transcendental entire functions and in this an application of (1) (that is, the insertion of linear factors) is usually dismissed as a triviality. We wish to point out that this may not be as trivial as it seems; for example, a polynomial may not commute with any other polynomial (except its own iterates), but for a suitable linear polynomial

ℓ, $\ell \circ p_j$ may commute with many other polynomials (for example, when $\ell \circ p$ is a Tchebychev polynomial). As no nontrivial example of commuting transcendental entire functions seems to be known this is not at present an issue, but if an example were to be found, the insertion of linear factors would have this deeper implication. The same is true of the rule (3), and in this case there are analogous examples for transcendental entire functions, for example $e^z \circ (e^z + z) = ze^z \circ e^z$ but here e^z is not prime.

References

1. BARTON, D.R. and ZIPPEL, R., Polynomial decomposition algorithms, *J. Symb. Comp.*, 1 (1985), 159-168.
2. HORWITZ, A.L., Even compositions of entire functions and related matters, *J. Austral. Math. Soc.* (Ser. A), 63 (1997), 225-237.
3. HORWITZ, A.L. and RUBEL, L.A., When is the composition of two power series even?, *J. Austral. Math. Soc.* (Ser. A), 56 (1994), 415-420.
4. KOZEN, D. and LANDAU, S., Polynomial decomposition algorithms, *J. Symb. Comp.*, 7 (1989), 445-456.
5. RITT, J.F., Equivalent rational substitutions, *Trans. Amer. Math. Soc.*, 26 (1924), 221-229.
6. RITT, J.F., Prime and composite polynomials, *Trans. Amer. Math. Soc.*, 23 (1926), 51-66.
7. VON ZUR GATHEN, J., Functional decomposition of polynomials : the tame case. *J. Symb. Comp.*, 9 (1990), 281-299.