UNIQUE SUBWORDS IN NONPERIODIC WORDS

C. M. WEINBAUM

(Communicated by Warren J. Wong)

Abstract. Let $A, D$ be words over some alphabet. $D$ has position $p$ in the cyclic word $A$ if the cyclic permutation of $A$ which begins with the $p$th letter of $A$ has an initial subword equal to $D$. It is proved that every nonperiodic word $A$ of length $> 1$ has a cyclic permutation which is a product $BC$ for some nonempty subwords $B, C$ having unique positions in the cyclic word $A$.

1. Introduction

The result in the Abstract can be viewed as an assertion about a group $\langle a, b, \ldots ; A = 1 \rangle$ with a single nonperiodic defining relator $A$ involving 2 or more generators $a, b, \ldots$ but none of the inverses of $a, b, \ldots$. The result may be useful for investigating the group's Cayley graph. (See [1, p. 57] for a definition.) We conjecture that a general result holds for any presentation $\langle a, b, \ldots ; A = 1 \rangle$ whenever $A$ is nonperiodic with length $> 1$.

Conjecture. Every nonperiodic word $A$ of length $> 1$ has a cyclic permutation which is a product $BC$ where nonempty words $B, C$ have unique positions in the cyclic word $A$ and $B, C$ have no positions in the cyclic word formed by the inverse of $A$.

For convenient exposition, finite sequences instead of words will be used to introduce terminology and state and prove the main result in Theorem 1.

2. Terminology for sequences

A finite nonempty sequence of objects is referred to briefly as a sequence. A well-defined relation of equality between objects is assumed. The length $n$ of a sequence $S$ is denoted by $|S|$. A sequence $T$ is an initial subsequence of $S$ if $T$ consists of the first $|T|$ terms of $S$. An extreme term of $S$ is either the first or last term of $S$. There is just one extreme term if $|S| = 1$. $S(j)$ denotes the...
jth term of \( S \). \([1, n]\) denotes the set of allowable values of \( j \), namely, the set of integers from 1 to \( n \), inclusive. An integer \( p \) in \([1, n]\) is a position for \( S \).

Let \( p \) be a position for a sequence \( S \). Let \( (Sp) \) denote the cyclic permutation of \( S \) whose first term is the \( p \)th term of \( S \). Then a sequence \( T \) is a segment of \( S \), having position \( p \), if \( T \) is an initial subsequence of \( (Sp) \). The product \( UV \) of sequences \( U, V \) is a sequence consisting of the terms of \( U \) followed by the terms of \( V \). The finite product \( P \), formed from a sequence \( S \) of sequences, is defined similarly. \( S \) is referred to as a factorization of \( P \); each term in \( S \) is a factor.

A sequence \( S \) is periodic if, for some integer \( k > 1 \), \( S \) is the product of \( k \) copies of some sequence \( X \). The period of \( S \), denoted by \( S \), is equal to \( |X| \) when \( k \) is as large as possible.

3. Terminology for run sequences

Choose an object \( z \) which may or may not be equal to some term in a given sequence \( A \). The run sequence \( rA \), based on \( z \), is the factorization of \( A \), with as few factors as possible, such that, for each factor, either all or none of the terms in that factor are equal to \( z \). A run is a term in \( rA \). If \( p \) is a position for \( A \) and \( A(p) \) is the first term in some run, then \( p \) is a starting position for \( A \). The sequence of starting positions for \( A \) is denoted by \( sA \). The sequences of odd- and even-numbered terms of \( rA \) are denoted by \( orA \), \( erA \), respectively. The expressions \( sA(j), rA(j), orA(j), erA(j) \) denote the \( j \)th terms of \( sA \), \( rA \), \( orA \), \( erA \), respectively.

4. Preliminaries

Lemma 1. Let sequence \( B \) be a segment of a sequence \( A \). Assume unequal extreme terms for \( A \). Let \( rB, rA \) be the run sequences based on \( A(1) \) so length of \( rA \) is \( 2m \) for some positive integer \( m \). Denote the sequences of odd-numbered terms of \( rB, rA \) by \( orB, orA \), respectively. Assume

(1) each extreme term of \( B \) is equal to \( A(1) \),
(2) each extreme term of \( rB \) is a sequence of length \( oMAX \) where \( oMAX \) is the maximum of the lengths of the sequences which appear as terms in \( orA \),
(3) \( orB \) has a unique position in \( orA \).

Then \( B \) has a unique position in \( A \).

Proof. Suppose \( B \) has position \( b \) in \( A \). Conditions (1), (2) imply that the extreme terms of \( rB \) are runs in \( rA \). Hence, \( rB \) is a segment of \( rA \) with odd position \( 2k - 1 \) in \( rA \) for some integer \( k \) in \([1, m]\). Thus \( orB \) has position \( k \) in \( orA \). But \( k \) is unique by (3). Hence \( rB, B \) have unique positions \( 2k - 1, sA(2k - 1) \) in \( rA, A \) respectively. (See §3 for definition of \( sA \).) So \( b = sA(2k - 1) \) and \( b \) is unique.

Lemma 2. Let sequence \( C \) be a segment of a sequence \( A \). Assume unequal extreme terms for \( A \). Assume \( C \) has a factorization given by \( C = XWY \). Let \( rW, rA \) be the run sequences based on \( A(1) \) so length of \( rA \) is \( 2m \) for some
positive integer \( m \). Denote the sequences of odd-numbered terms of \( rW, rA \) by \( \text{or}W, \text{or}A \), respectively. Assume that

1. sequences \( X, Y \) contain no terms equal to \( A(1) \),
2. each extreme term of \( W \) is equal to \( A(1) \),
3. \( \text{or}W \) has a unique position in \( \text{or}A \).

Then \( C \) has a unique position in \( A \).

Proof. Suppose \( C \) has position \( c \) in \( A \). Conditions (1), (2) imply \( rW \) is a segment of \( rA \). So \( rW \) has odd position \( 2k - 1 \) in \( rA \) for some integer \( k \) in \([1, m]\). Thus \( \text{or}W \) has position \( k \) in \( \text{or}A \). But \( k \) is unique by (3). Hence, \( rW, W \) have unique positions \( 2k - 1, sA(2k - 1) \) in \( rA, A \), respectively. (See §3 for definition of \( sA \).) So \( c \) is congruent, modulo \( |A| \), to \( sA(2k - 1) - |X| \). Hence, \( c \) is unique.

**Lemma 3.** Let a sequence \( U \) have positions \( c, d \) in a periodic sequence \( V \) with period \( \overline{V} \). If \( |U| = \overline{V} - 1 \), then \( c, d \) are congruent modulo \( \overline{V} \).

**Lemma 4.** Let \( S, T \) be sequences with \( S \) a segment (of \( T \)) having an odd position in \( T \). Assume

1. \( T \) has even length (\( |T| = 2m \) with \( m \) a positive integer.),
2. the odd and even numbered subsequences \( oT, eT \) of \( T \) have no common terms,
3. \( oT, eT \) are periodic with periods \( oT, eT \), satisfying \( 1 < eT < oT \),
4. \( m \) is the least common multiple of \( oT \) and \( eT \),
5. \( |S| = 2oT - 3 \).

Then \( S \) has a unique position in \( T \).

Proof. Suppose \( S \) has positions \( p, q \) in \( T \). Then \( oS \) has positions \( (1+p)/2, (1+q)/2 \) in \( oT \). Also \( eS \) has positions \( (1+p)/2, (1+q)/2 \) in \( eT \) and so does \( eW \) for each initial subsequence \( W \) of \( S \). Define such a \( W \) by requiring \( |W| = 2eT - 2 \). Then \( |W| \) does not exceed \( |S| \) because of (3), (5). Note that \( |eW| = eT - 1 \). Apply Lemma 3 to \( V = oT, U = oS \) and to \( V = eT, U = eW \) to find \( (1+p)/2, (1+q)/2 \) congruent modulo \( oT \), modulo \( eT \) and hence modulo \( m \), using (4). Hence, \( 1 + p, 1 + q \) are congruent modulo \( 2m \), i.e. modulo \( |T| \). So \( p, q \) are positions for \( T \) and are congruent modulo \( |T| \). Thus \( p = q \).

**5. Main result**

**Theorem 1.** Each finite nonperiodic sequence \( A \) of length \( n > 1 \) has a cyclic permutation which can be expressed as a product \( BC \) where each of the sequences \( B, C \) has a unique position in \( A \). For any integer \( p \) in \([1, n]\), the sequences \( B, C \) can be chosen so that each extreme term of \( B \) is equal to \( A(p) \) and each extreme term of \( C \) is not equal to \( A(p) \).

Proof. Use induction on \( n \). The theorem is obvious when \( n = 2 \). Unequal terms \( A(1), A(n) \) and \( p = 1 \) can be assumed. Each run sequence defined in the proof, including \( rA \), is based on the object \( A(1) \). Denote the odd-
and even-numbered subsequences of \( rA \) by \( orA \), \( erA \), respectively. Then \( |rA| = 2m \), \( |orA| = |erA| = m \) for some positive integer \( m \). In each of five cases, definitions of \( B \), \( C \) assure that the extreme terms of \( B \), \( C \) satisfy the theorem.

Case 1. \( m = 1 \). In this case, use \( B = rA(1) \) and \( C = rA(2) \).

Case 2. \( orA \) is not periodic and \( m > 1 \).

Define \( o MAX \) to be the maximum of \( |orA(j)| \) for \( j \) in \([1, m]\). Pick an integer \( q \) in \([1, m]\) so that \( |orA(q)| = o MAX \). By the induction assumption we can apply Theorem 1 to the sequence \( orA \). Thus there exist sequences \( P \), \( Q \) such that \( PQ \) is a cyclic permutation of \( orA \), the positions for \( P \), \( Q \) in \( orA \) are unique, and each extreme term of \( P \) is equal to \( orA(q) \). Now assume that \( orA = PQ \) because, if not, the remainder of the proof for Case 2 can be applied to the cyclic permutation \( A' \) of \( A \) for which \( orA' = PQ \). Then the sequences \( B \), \( C \) found for \( A' \) will suffice for \( A \).

Define \( B \) to be the product of the first \( 2|P| - 1 \) terms in \( rA \). Then \( rB \), \( orB \) are initial subsequences of \( rA \), \( orA \), respectively and \( orB = P \). Apply Lemma 1 to conclude that \( B \) has a unique position in \( A \).

Define \( C \) by \( A = BC \). Then each extreme term of \( C \) is not equal to \( A(1) \) because each extreme term of \( rB \) has length \( o MAX \). Thus \( rA = rBrC \). So \( C = XWY \) where \( X, Y \) are the extreme terms of \( rC \) and \( orW = Q \). Apply Lemma 2 to conclude that \( C \) has a unique position in \( A \).

Case 3. \( erA \) is not periodic and \( m > 1 \).

Let \( A = XY \) where \( X = rA(1) \). Let \( D = YX \). Then \( orD = erA \). So Case 2 applies to \( D \).

Case 4. \( orA \) and \( erA \) are periodic (hence, \( m > 1 \)) and the periods \( orA \), \( erA \), satisfy \( orA < erA \).

In this case, \( m \) equals the least common multiple of \( orA \) and \( erA \) else \( A \) would be periodic. Define \( B \) to be the product of the first \( 2erA - 1 \) terms in \( rA \). Define \( C \) by \( A = BC \). Then \( B = XYZ \) where \( X, Z \) are the first, last runs of \( B \) and \( Y \) is the product of the middle \( 2erA - 3 \) runs of \( B \). Thus \( rY \) is a segment with position 2 in \( rA \), with \( |rY| = 2erA - 3 \). Using \( D = YZCX \), we find that \( rY \) is an initial subsequence of \( rD \) and \( orD = erA \), \( erD = orA \). Apply Lemma 4 using \( T = rD \), \( S = rY \). So \( rY \) has a unique position in \( rD \); hence \( rY \) has a unique position, namely 2, in \( rA \).

Now suppose \( B \) has position \( b \) in \( A \). Let \( i, j \) denote the positions for \( A \) which are congruent modulo \( |A| \) to \( b + |X| - 1 \) and \( b + |X| \), respectively. Then the last term of \( X \) and the first term of \( Y \) have positions \( i, j \), respectively, in \( A \). So \( Y \) has position \( j \) in \( A \). Since all the terms in \( X \) and none of the extreme terms in \( Y \) are equal to \( A(1) \), \( j \) is a starting position for \( A \), appearing as an even-numbered term in \( sA \) (which is defined in §3). Thus \( j = sA(2k) \) for some integer \( k \) in \([1, m]\). Therefore \( rY \) has position \( 2k \) in \( rA \), but this position is unique and equal to 2. So \( j = sA(2k) = sA(2) = 1 + |X| \). Hence
1 + |X|, \ b + |X| \ are congruent modulo |A|. Thus b = 1 and B has a unique position in A.

Now apply the reasoning about B to another segment H (of A), also having an initial subsequence equal to an odd-numbered run of A. Define H to be the product of 2erA - 1 terms in rA, beginning with the term rA(2erA + 1). Deduce that H has a unique position in A. Define C by A = BC. Then C has a unique position in A because so does H, a subsequence of C consisting of consecutive terms of C.

Case 5. orA and erA are periodic (hence, m > 1) and periods orA, erA, satisfy erA < orA.

Let A = GH where G is the first run in rA. Define A' = HG so that the last term of A' equals A(1) and A' has unequal extreme terms. Then orA' = erA and erA' = orA. Now apply the reasoning in Case 4 to A'.

Define B' to be the product of the first 2erA' - 1 terms in rA'. Define C' by A' = B'C'. Each extreme term of B' equals A'(n) and each extreme term of C' equals A(1). As in Case 4, B', C' have unique positions in A'. Now use B = C', C = B' so that BC is a cyclic permutation of B'C' which is a cyclic permutation of A. This completes the proof.

**References**


**Umecorp, 45 San Clemente Drive, Corte Madera, California 94925**