There exist binary circular $5/2^+$ power free words of every length

Ali Aberkane & James D. Currie*
Department of Mathematics and Statistics
University of Winnipeg
Winnipeg, Manitoba R3B 2E9, Canada
e-mail: aberkane@iml.univ-mrs.fr,
currie@uwinnipeg.ca

Submitted: Oct 31, 2003; Accepted: Jan 5, 2003; Published: Jan 23, 2004
MR Subject Classification: 68R15

Abstract

We show that there exist binary circular $5/2^+$ power free words of every length.
Keywords: Combinatorics on words, Dejean's conjecture, Thue-Morse word

1 Introduction

The word *alfalfa* consists of the segment *alfa* overlapped with itself. Alternatively, we may view *alfalfa* as *alf*, taken $2\frac{1}{3}$ times; we write *alfalfa* $=$ *alf* $^{7/3}$.

Let w be a word, $w = w_1w_2 \dots w_n$ where the w_i are letters. We say that w is **periodic** if for some integer $p \leq n$ we have $w_i = w_{i+p}$, $i = 1, 2, \dots, n-p$. We call p a **period** of w. Thus by convention, length n of w is always a period. Let k be a rational number. If p is a period of w, and $|w| = kp$, then we say that w is a k **power**. For example, every word is 1 power. A k^+ **power** is a word which is an r power for some $r > k$. A word is k^+ **power free** if none of its subwords is a k^+ power. A 2 power is called a **square**, while a 2^+ power is called an **overlap**.

Thue showed that there are infinite sequences over $\{a, b\}$ not containing any overlaps, and infinite sequences over $\{a, b, c\}$ not containing any squares [7]. As well as studying sequences, Thue studied necklaces or circular words.

Word v is a **conjugate** of word w if there are words x and y such that $w = xy$ and $v = yx$. Let w be a word. The **circular word** w is the set consisting of w and all of its conjugates. We say that **circular word** w **is** k^+ **power free** if all of its elements are k^+

*The author's research was supported by an NSERC operating grant.

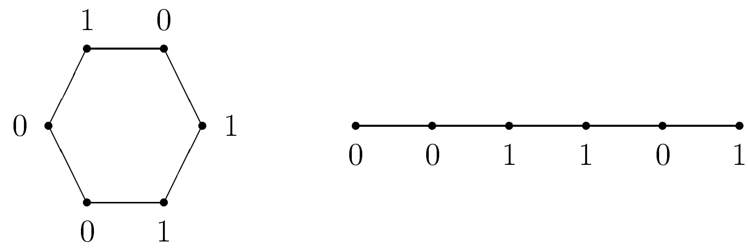

Figure 1: A 2+ free circular word.

power free; that is, all the conjugates of the 'ordinary word' w are k^+ power free. Thue proved that overlap-free binary circular words of length n exist exactly when n is of the form 2^m or 3×2^m.

Example 1 The set of conjugates of word 001101 is

$$\{001101, 011010, 110100, 101001, 010011, 100110\}.$$

Each of these is 2^+ power free, so that 001101 is a circular 2^+ power free word. (See Figure 1.) On the other hand, 0101101 is 3^+ power free, but its conjugate 1010101 is a 7/2 power. Thus 0101101 is not a circular 3^+ power free word.

Dejean [3] generalized Thue's work on repetitions to fractional exponents. Define the **repetitive threshold function** by

$$RT(n) = \sup\{k : x^k \text{ is unavoidable on } n \text{ letters}\}.$$

Dejean conjectured that

$$RT(n) = \begin{cases} 2, & n = 2 \\ 7/4, & n = 3 \\ 7/5, & n = 4 \\ n/(n-1), & n > 4 \end{cases}$$

We see that both Thue and Dejean studied the question of whether infinite sequences avoiding k powers exist over a given alphabet. In the case of 'linear words', i.e. sequences, this question has several equivalent formulations:

- Over an n-letter alphabet, are there arbitrarily long k power free words?
- Over an n-letter alphabet, are there k power free words? of every length $n > N_0$, some N_0?
- Over an n-letter alphabet, are there k power free words? of every length?

These formulations are equivalent, since the linear k power free words are closed under taking subwords. For circular words, these formulations become three distinct questions. As mentioned above, Thue showed that there are arbitrarily long binary circular words avoiding 2+ powers, but only for lengths of the form 2^m or 3×2^m. It was recently shown [1] that there are ternary square-free circular words of length n for $n \geq 18$. (Such words do not exist for for $n = 5$, for example.) On the other hand, there are binary cube-free circular words of every length [2]; in fact, such words can be found in the Thue-Morse sequence [5].

The three formulations give three possible generalizations of Dejean's work. We consider what seems to us the most natural of these

Let n be a positive integer, and k a rational number. Let $L(n, k)$ be the set of positive integers m such that no circular k power free word over n letters has length m. Every non-empty word is a 1 power; therefore, $L(n, 1)$ is always the set of positive integers. In particular, $L(n, 1)$ is non-empty. Define

$$CRT(n) = \sup\{k : L(n, k) \text{ is non-empty}\}.$$

We demonstrate that $CRT(2) = 5/2$. Thus, we prove the following:

Main Theorem: *Let n be a natural number. There is a circular binary word of length n simultaneously avoiding k powers for every rational $k > 5/2$.*

One quickly checks that every circular binary word of length 5 contains either a cube or a 5/2 power. Combining this observation with the theorem, one has $CRT(2) = 5/2$, as claimed. We have found 5/2+ free circular words of lengths up to 200 in the Thue-Morse word, leading us to make the following conjecture:

Conjecture 2 Let n be a natural number. The Thue-Morse sequence contains a subword of length n which, as a circular word, simultaneously avoids x^k for every rational $k > 5/2$.

2 A few properties of the Thue-Morse word

The Thue-Morse word t is defined to be $t = h^\omega(0) = \lim_{n\to\infty} h^n(0)$, where $h : \{0,1\}^* \to \{0,1\}^*$ is the substitution generated by $h(0) = 01$, $h(1) = 10$. Thus

$$t = 0110100110010110100101100110 1001 \cdots$$

The Thue-Morse word has been extensively studied. (See [4, 6, 7] for example.) We use the following facts about it:

1. Word t is 2^+ power free.
2. If w is a subword of t then so is $\overline{w}$. (The set of subwords of t is closed under taking binary complements.)
3. None of 00100, 01010, 10101 or 11011 is a subword of t.

Lemma 3 *Let $k \geq 2$ be a positive integer. Then t contains subwords of length k of the form $0v1$ and of the form $0v0$.*

Proof: Suppose that t has no subword $0v1$ of length k. Then any subword of t of length k which begins with a 0 must end with a 0. Since t is closed under binary complements, any subword of t of length k which begins with a 1 must end with a 1. This means that t is periodic with period $k-1$. This is absurd, since t is 2^+ power free. A similar contradiction arises if we assume that t has no subword $0v0$ of length k; in this case t would be periodic with period $2k-2$.□

Lemma 4 *Let $k \geq 6$ be a positive integer. Then t contains a subword of length k of the form $01v01$ and a subword of length k of the form $01v10$.*

Proof: If k is even, let $k = 2r$. We have $r = k/2 \geq 3$, so that t contains a word $u = 0v0$ of length r by the last lemma. Word $h(u) = 01h(v)01$, a word of the required form of length k.

If k is an odd integer, $k \geq 7$, we can write k as $8r-9$, $8r-7$, $8r-5$ or $8r-3$ for some $r \geq 2$. Let $u = 0v0$ be a word of length r in t. The word

$$h^3(u) = 011\underline{01}001h^3(u)\underline{01}1010\underline{01}$$

contains words $01v01$ of lengths $8r-9$ (including the first and second underlined 01's) and $8r-3$ (including the first and third underlined 01's.)

Let $z = 0v1$ be a word of length r in t.The word

$$h^3(z) = 011\underline{01}001h^3(u)10\underline{0101}10$$

contains words $01v01$ of lengths $8r-7$ (including the first and second underlined 01's) and $8r-5$ (including the first and third underlined 01's.)

The proof for $01v10$ is analogous.□

Applying h^2 to the words of the previous lemma gives the following corollary.

Corollary 5 *Let $k \geq 6$ be a positive integer. Then t contains subwords of length $4k$ of the form $= 01101001v01101001$ and of the form $01101001v10010110$.*

3 Circular $5/2^+$ power free words

Consider the words

- $f_0 = 00100$
- $f_1 = 01010$
- $f_2 = 10101$

- $f_3 = 11011$

None of the f_i appears in the Thue-Morse word t. (The 'f' is for 'forbidden'.) Note that f_i is the binary complement of f_{3-i}, $i = 0, 1$. For certain i and j we introduce words $b_{i,j}$ form 'buffers' between f_i and f_j. The words $b_{i,j}$ can be any subwords of the Thue-Morse word t with $|b_{i,j}| \geq 32$, and of the following forms:

- $b_{0,0} = 1101001v1001011$
- $b_{1,1} = 01101001v10010110$
- $b_{3,0} = 0010110v1001011$
- $b_{0,3} = 1101001v0110100$
- $b_{1,2} = 01101001v01101001$
- $b_{2,1} = 10010110v10010110.$

Again, there is symmetry; interchanging subscripts i and $3-i$ simply produces a binary complement. The condition that these words lie in t implies that each v will have either 0110 or 1001 as a prefix. These words are obtained from the words of Corollary 5, possibly taking the binary complement, and/or deleting the first and last letters. We see then that words $b_{0,0}$, $b_{3,0}$, $b_{0,3}$ exist for every length $4k-2$, $k \geq 9$. (We use $k \geq 9$ rather than $k \geq 6$ because we want $|b_{i,j}| \geq 32$.) Words $b_{1,1}$, $b_{1,2}$, $b_{1,2}$ exist for every length $4k-4$, $k \geq 9$.

Let w be a circular word of one of the forms

$$\begin{array}{l} b_{0,0}f_0 \\ b_{1,1}f_1 \\ b_{3,0}f_0f_3 \\ b_{1,2}f_2b_{2,1}f_1 \end{array} \tag{1}$$

By controlling the lengths of the $b_{i,j}$, word w can be chosen to have length $4k_1+3$, $4k_1+1$, $4(k_1)+8$ or $4(k_1+k_2)-8+10$ for any $k_1, k_2 \geq 9$. In particular, word w can have any length $n \geq 74$. We claim that w avoids all x^k with $k > 5/2$. The proof begins with the following lemma:

Lemma 6 *No word of the form $ab_{i,j}c$ with $|a|, |c| \leq 4$ is a k power for rational $k > 5/2$.*

Proof: Suppose $ab_{i,j}c$ is a k power for $k > 5/2$, where $|a|, |c| \leq 4$. This means that $ab_{i,j}c$ is periodic with some period p, $|ab_{i,j}c| > 5p/2$. Its subword $b_{i,j}$ must also then have period p. Since $b_{i,j}$ is a subword of t, this means that $|b_{i,j}| \leq 2p$. In total then, $8 \geq |a| + |c| = |ab_{i,j}c| - |b_{i,j}| > 5p/2 - 2p = p/2$, so that $16 > p$. However, then $32 \leq |b_{i,j}| \leq 2p \leq 2 \times 15 = 30$. This is a contradiction.□

Lemma 7 *Suppose that a word of the form $s\beta$ is a k power for rational $k > 5/2$, where, for some i and j, word f_i has suffix s, $|s| \leq 4$ and $b_{i,j}$ has β as a prefix. Let $s\beta$ have period $p < 2|s\beta|/5$. Then $p \leq 7$.*

Proof: The word β has period p, but is a subword of t. Thus, $|\beta| \leq 2p$. Now, $4 \geq |s| = |s\beta| - |\beta| > 5p/2 - 2p = p/2$. We conclude that $7 \geq p$.$\Box$

Lemma 8 *Consider a word of the form $s\beta$ where, for some i and j, β is a prefix of $b_{i,j}$, s is a suffix of f_i, $|s| \leq 4$. Then for rational $k > 5/2$, $s\beta$ is not a k power.*

Proof: By symmetry, it suffices to prove the result where i is 0 or 1.

Case 1: We suppose $i = 0$.

Word s will be a suffix of 0100. Let $\pi_1 = 1101001\ 0110$ and let $\pi_2 = 1101001\ 1001$. (The spaces are for clarity; they highlight the two possible prefixes of v in $b_{i,j}$.) By the construction of $b_{0,0}$ and $b_{0,3}$, one of π_1, π_2 is a prefix of $b_{i,j}$. It follows that either β is a prefix of one of the π_k, or one of the π_k is a prefix of β.

Let $s\beta$ have period p, $|s\beta| > 5p/2$. By Lemma 7, $p \leq 7$. If π_k is a prefix of β, then $s\pi_k$ has period p. On the other hand, if β is a prefix of π_k, then $s\pi_k$ has a prefix $s\beta$, $|s\beta| > 5p/2$. Let q be the maximal prefix of $s\pi_k$ with period p. For each choice $p = 1, 2, \ldots, 7$, and for each possibility $k = 1, 2$, we show two things:

1. Word q is a proper prefix of $s\pi_k$. This eliminates the case where π_k is a prefix of β.

2. We have $|q| \leq 5p/2$. This eliminates the case where β is a prefix of π_k. We thus obtain a contradiction.

As an example, suppose $p = 6$. In $s\pi_1 = s1\mathbf{1}01001\ \mathbf{0}110$, the letters in bold-face differ. This means that prefix q of period 6 is a prefix of $s1101001$, which has length $|s| + 7 \leq 11 \leq 5p/2 = 5 \times 6/2 = 15$. Again, in $s\pi_2 = s110\mathbf{1}001\ 10\mathbf{0}1$, the letters in bold-face differ. Any prefix of $s\pi_2$ of period 6 is thus a prefix of $s110100110$, which has length at most 14.

The following table bounds $|q|$ in the various cases. The pairs of bold-face letters certify the given values.

p	s	π_i		$\lvert q\rvert$		$\lvert q\rvert/p$
1	(0)**0**	**1**101001v	$\leq$	2	$\leq$	2
	(0)**10**0	1101001v	$\leq$	2	$\leq$	2
2	(010)**0**	1**1**01001v	$\leq$	5	$\leq$	5/2
3	0	1**1**01**0**01v		5		5/3
	(01)**0**0	1**1**01001v	$\leq$	5	$\leq$	5/3
4	(010)**0**	110**1**001v	$\leq$	7	$\leq$	7/4
5	(010)0	**1**1010**0**1v	$\leq$	9	$\leq$	9/5
6	(010)0	1**1**01001 **0**110	$\leq$	11	$\leq$	11/6
	(010)0	110**1**001 10**0**1	$\leq$	14	$\leq$	7/3
7	(010)**0**	110100**1**v	$\leq$	10	$\leq$	10/7

The parentheses abbreviate rows of the table. For example, cases $s=0$ and $s=00$ are together in the first row of the table. The bold-faced pair will work whether $s=0$ or $s=00$. We have q a prefix of s, whence $|q|\le 2$. Similarly, when $p=5$, one pair works for all values of s.

Case 2: We suppose $i=1$.

Let $\rho_1 = 01101001\,0110$, $\rho_2 = 01101001\,1001$. In analogy to the previous case, the following table completes the proof:

p	s	ρ_i		$\lvert q\rvert$		$\lvert q\rvert/p$
1	**0**	**01**101001v		2		2
	10	01101001v		1		1
	010	01101001v		1		1
	1010	01101001v		1		1
2	**0**	0**1**101001v		2		1
	(10)**10**	**0**1101001v	$\le$	4	$\le$	2
3	(101)**0**	01**1**01001v	$\le$	6	$\le$	2
4	(1010)	**0**110**1**001v	$\le$	8	$\le$	2
5	(101)**0**	0110**1**001v	$\le$	8	$\le$	8/5
6	(101)0	01**1**01001 **0**110	$\le$	12	$\le$	2
	(101)0	0110**1**001 10**0**1	$\le$	14	$\le$	7/3
7	(101)0	**0**110100**1**v	$\le$	11	$\le$	11/7

Evidently, one could also verify this lemma via computer.□

Lemma 9 *Consider a word of the form βr where, for some i and j, β is a suffix of $b_{i,j}$, r is a prefix of f_j, $|r|\le 4$. Then for rational $k>5/2$, βr is not a k power.*

Proof: This assertion follows from the last by symmetry.□

Corollary 10 *Let w be a word of form 1, and let w contain a k power z, some rational $k>5/2$. Then z contains some f_i, $i=0,1,2$ or 3.*

Proof: Word z is an ordinary subword of some conjugate of w. The conjugates of w have one of the forms $b''f_ib'$, $f''b_{i,i}f'$, $b''f_jb_{j,i}f_ib'$, $b''f_0f_3b'$ or $f''b_{i,j}f_jb_{j,i}f'$ where $f_i=f'f''$ and $b_{i,j}=b'b''$, some i and j. We know that z cannot be a subword of any $b_{i,j}$, since t is 2^+ power free. If z does not contain any f_i therefore, then z has one of the forms $f''b_{i,j}f'$, $f''b'$ or $b''f'$, where $|f'|$, $|f''|\le 4$. These possibilities are ruled out by Lemmas 6, 8 and 9 respectively.□

Lemma 11 *Suppose z is a periodic word with period p and $|z|>5p/2$. Let x be a subword of z with $|x|\le p/2$. Then z contains a subword xyx for some y.*

Proof: Let ax be a prefix of z with a as short as possible. As z is periodic, $|a| < p$. This implies that $|ax| = |a| + |x| < p + p/2 = 3p/2$. It follows that $|ax| + p < 5p/2 < |z|$, and the result follows.□

Remark 12 The words f_i, $i = 0, \ldots, 3$ never appear in t. It follows that each of these words appears at most once in any conjugate of w.

Lemma 13 *Let w be a word of form 1, and let w contain a k power z, some rational $k > 5/2$. Let z have period p. Then $p \leq 9$.*

Proof: By Remark 12, z contains each of the f_i at most once. By Corollary 10, z contains one of the f_i exactly once. Thus z contains some word x exactly once, where $|x| = 5$. By the contrapositive of Lemma 11, $p < 2|x| = 10$.□

Theorem 14 *Let w be a word of form 1. Then word w is $5/2^+$ power free.*

Proof: Suppose for the sake of getting a contradiction that a conjugate of w contains a k power z, some $k > 5/2$. Let z have period p, $|z| = kp$. By the last lemma, $p \leq 9$. Without loss of generality, shortening z if necessary, suppose that $|z| = \lceil 5p/2 \rceil$. This implies that $|z| \leq \lceil 45/2 \rceil = 23$.

By Remark 12, z contains f_i for some i. Since $|z| \leq 23$, we have one of two cases:

Case A: We can write $z = af_ic$ where c is a prefix of $b_{i,j}$ for some j, and $b_{m,i}$ has suffix a for some m.

Case B: We can write $z = af_0f_3c$ where c and a are prefix and suffix respectively of $b_{3,0}$.

Proof in Case A: Using symmetry, we may assume that $i = 0$ or $i = 1$.

Case A1: We suppose $i = 0$.

As in the proof of Lemma 8, we take $\pi_1 = 1101001\,0110$, $\pi_2 = 1101001\,1001$. Also, let $\nu_1 = 0110\,1001011$ and let $\nu_2 = 1001\,1001011$. One of the words π_k must be a prefix of c, or vice versa. Similarly, either a is a suffix of one of the ν_k, or one of the ν_k is a suffix of a.

Word f_0 does not have period 1 or 2. Therefore, $p \geq 3$. In the case where $p = 3$, f_0 sits in $\nu_k f_0 \pi_m$ in the context 0**1**1 0**0**1$\underline{0}$0 1$\underline{1}$0. As in the proof of Lemma 8, the bold-faced pair limit the possible extent of z on the left. In addition, the underlined pair limit z on the right. In total, $|z| \leq |1001001| = 7 \leq 5/2 \times 3$. Thus $p = 3$ gives a contradiction. Similar contradictions are obtained for $p = 4$ to 9, as set out in the following table:

p	$\nu_k f_0 \pi_m$		$\lvert z\rvert$		$\lvert z\rvert/p$
4	$\cdots\mathbf{1}\ 0\underline{0}1\mathbf{0}0\ \underline{1}\cdots$	$\leq$	5	$\leq$	5/4
5	$\cdots\mathbf{1}\ \underline{0}010\mathbf{0}\ \underline{1}\cdots$	$\leq$	5	$\leq$	1
6	$\cdots\mathbf{1}1\ \underline{0}010\mathbf{0}\ 1\underline{1}\cdots$	$\leq$	7	$\leq$	7/6
7	$\cdots\mathbf{1}011\ 0\underline{0}1\mathbf{0}0\ 110\underline{1}\cdots$	$\leq$	11	$\leq$	11/7
8	$\cdots\mathbf{1}011\ \underline{0}010\mathbf{0}\ 110\underline{1}\cdots$	$\leq$	11	$\leq$	11/8
9	$01\mathbf{1}01001011\ \mathbf{0}010\underline{0}\ 11010010\underline{1}10$	$\leq$	21	$\leq$	7/3
9	$01\mathbf{1}01001011\ \mathbf{0}01\underline{0}0\ 1101001\underline{1}001$	$\leq$	20	$\leq$	20/9
9	$100\mathbf{1}1001011\ 0\mathbf{0}10\underline{0}\ 11010010\underline{1}10$	$\leq$	20	$\leq$	20/9
9	$100\mathbf{1}1001011\ 0\mathbf{0}1\underline{0}0\ 1101001\underline{1}001$	$\leq$	19	$\leq$	19/9

Case A2: We suppose $i = 1$.

This time we take $\rho = 01101001$. Let $\sigma 10010110$. Word f_1 does not have period 1 or 3, so the proof is finished as set out in the following table:

p	$\sigma f_1 \rho$		$\lvert z\rvert$		$\lvert z\rvert/p$
2	$\cdots\mathbf{0}\ 0\mathbf{1}0\underline{1}0\ \underline{0}\cdots$	$\leq$	5	$\leq$	5/2
4	$\cdots\mathbf{0}\ 0\underline{1}0\mathbf{1}0\ \underline{0}\cdots$	$\leq$	5	$\leq$	5/4
5	$\cdots\mathbf{1}10\ 01\underline{\mathbf{0}}10\ 01\underline{1}\cdots$	$\leq$	9	$\leq$	9/5
6	$\cdots\mathbf{1}0\ \underline{0}101\mathbf{0}\ 0\underline{1}\cdots$	$\leq$	7	$\leq$	7/6
7	$\cdots\mathbf{1}10\ \underline{0}101\mathbf{0}\ 01\underline{1}\cdots$	$\leq$	9	$\leq$	9/7
8	$1\mathbf{0}010110\ 0\mathbf{1}0\underline{1}0\ 011010\underline{0}1$	$\leq$	17	$\leq$	17/8
9	$\cdots\mathbf{1}0110\ \underline{0}101\mathbf{0}\ 0110\underline{1}\cdots$	$\leq$	13	$\leq$	13/9

Proof in Case B: This case cannot occur, since $f_0 f_3$ does not have period $p \leq 9$, as documented in the following table:

p	$f_0 f_3$
1	$0\mathbf{01}0011011$
2	$\mathbf{0}0\mathbf{1}0011011$
3	$001\mathbf{0}01\mathbf{1}011$
4	$0\mathbf{0}100\mathbf{1}1011$
5	$\mathbf{0}0100\mathbf{1}1011$
6	$\mathbf{0}01001\mathbf{1}011$
7	$0\mathbf{0}100110\mathbf{1}1$
8	$\mathbf{0}0100110\mathbf{1}1$
9	$\mathbf{0}01001101\mathbf{1}$

Main Theorem: *Let n be a natural number. There is a circular binary word of length n simultaneously avoiding k powers for every rational $k > 5/2$.*

Proof: One can find circular $5/2^+$ power free words up to length 73 in the Thue-Morse word t. On the other hand, word w can be made to have any length 74 or greater.$\square$

References

[1] James D. Currie, There are ternary circular square-free words of length n for $n \geq 18$, *Elec. J. Comb.* **9(1)** N10.

[2] James D. Currie & D. Sean Fitzpatrick, Circular words avoiding patterns, **Developments in Language Theory, - 6th International Conference**, DLT 2002, Kyoto, Japan, Lecture Notes in Computer Science, to appear (edited by Masami Ito and Masafumi Toyama) (2003) (Springer-Verlag).

[3] Françoise Dejean, Sur un théorème de Thue, J. Combin. Theory Ser. A **13** (1972), 90–99.

[4] Earl D. Fife, Binary sequences which contain no BBb, Trans. Amer. Math. Soc. **261** (1980), 115–136; MR **82a**:05034

[5] D. Sean Fitzpatrick, There are binary cube-free circular words of length n contained within the Thue-Morse word for all positive integers n. *Ars Combinatoria*. To Appear.

[6] Marston Morse & Gustav A. Hedlund, Symbolic dynamics I, II, Amer. J. Math. **60** (1938), 815–866; **62** (1940), 1–42; MR **1**, 123d.

[7] Axel Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana (1912), 1–67.

[8] A. Zimin, Blocking sets of terms, Mat. Sb. (N.S.) **119 (161)** (1982); Math. USSR Sbornik **47** (1984), 353–364.