Theoretical Computer Science 409 (2008) 432–437

Contents lists available at ScienceDirect

Theoretical Computer Science

journal homepage: www.elsevier.com/locate/tcs

Long binary patterns are Abelian 2-avoidable

James D. Currie *, Terry I. Visentin

Department of Mathematics and Statistics, University of Winnipeg, Canada

ARTICLE INFO

Article history:
Received 22 May 2007
Received in revised form 18 January 2008
Accepted 25 August 2008
Communicated by M. Crochemore

Keywords:
Combinatorics on words
Pattern avoidance
Non-repetitive sequences

ABSTRACT

We show that every long binary pattern is Abelian 2-avoidable.

1. Introduction

In Combinatorics on Words, the study of *words avoiding patterns* involves a class of decision problems with connections to Universal Algebra, and to the Post Correspondence Problem.

The first to study pattern avoidance problems was Thue. In his words [13,2]:

> "We consider the question whether, given u and w, there always exists[1] a nonerasing morphism $h : A^* \to B^*$ such that $h(u)$ is a factor of w".

He showed that when $u = xx$, there are infinitely many words $w \in \{1, 2, 3\}^*$ containing no factor of the form $h(xx)$, h non-erasing. The question of avoiding an arbitrary word u was taken up by workers in Universal Algebra in the late 1970's [1,14].

Let $p = p_1p_2 \cdots p_n$, where the p_i are letters. We say that w **encounters** p if w contains a factor $P_1P_2 \cdots P_n$ where the P_i are non-empty words with $P_i = P_j$ whenever $p_i = p_j$. In this case, $P_1P_2 \cdots P_n$ is called an **instance** of p. Otherwise, w **avoids** p and is p**-free**.

We refer to the word p which is to be encountered or avoided as a **pattern**. Pattern p is k**-avoidable** if there are infinitely many words over $\{1, 2, \ldots, k\}$ which avoid p. Pattern p is **avoidable** if it is k-avoidable for some k.

The problem of deciding whether a given pattern is avoidable was solved independently in [1] and [14]. The problem of deciding k-avoidability of a pattern remains open; many results are tabulated in [3] concerning k-avoidability of binary and ternary words, including a complete answer to the k-avoidability of binary words. In particular, every binary pattern of length 6 or more is 2-avoidable [12].

Erdös [7] proposed an Abelian (commutative) version of the problem solved by Thue: are there words of arbitrary length over a fixed finite alphabet not containing factors XX', where X' is obtained from X by rearranging letters? Problems of this flavour are mentioned by Zimin [14], who explicitly notes that his methods do not apply to the Abelian situation.

* Corresponding author.
E-mail addresses: j.currie@uwinnipeg.ca (J.D. Currie), t.visentin@uwinnipeg.ca (T.I. Visentin).

[1] Here A and B are finite alphabets, u a word over A, and w a word over B.

0304-3975/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.tcs.2008.08.039

Let p, q be words. We write $p \sim q$ if p and q are anagrams of each other; for example *posts* $\sim$ *stops*. Let $p = p_1 p_2 \cdots p_n$ where the p_i are letters. We say that w **encounters** p **in the Abelian sense** if w contains a factor $P_1 P_2 \cdots P_n$ where the P_i are non-empty words with $P_i \sim P_j$ whenever $p_i = p_j$. In this case, $P_1 P_2 \cdots P_n$ is called an Abelian **instance** of p. Otherwise w **avoids** p **in the Abelian sense** and is **Abelian** p**-free**.

Example 1.1. The word *protractor* $= p\ ro\ t\ rac\ t\ or$ encounters *abcba* in the Abelian sense.

Again, we refer to the word p which is to be encountered or avoided as a **pattern**. Pattern p is **Abelian** k**-avoidable** if there are infinitely many words over $\{1, 2, \ldots, k\}$ which avoid p in the Abelian sense. Pattern p is Abelian avoidable if it is Abelian k-avoidable for some k.

Let $a(n)$ be the smallest k such that x^n is k-avoidable. In a series of papers [8,11,6,9] it was shown that $a(4) = 2$, $a(3) = 3$, $a(2) = 4$. This last result was not established until 1992(!) reflecting the difficulty of the Abelian version of Thue's question.

Nevertheless, some progress has been made on Abelian pattern avoidance. In 2000 it was shown [4] that any pattern p on n letters with $|p| \geq 2^n$ is Abelian avoidable. (Such patterns were already known to be avoidable 'in the ordinary sense'.) Interestingly, the pattern $p =$ *abcabadabacba* was shown to be avoidable, but not Abelian avoidable.

An attack on Abelian k-avoidability began in [5], where a tool for the case $k = 2$ was introduced. In the present paper, we exploit and improve the tools used to present a finiteness result:

Theorem 1.2. *Binary patterns of length greater than* 118 *are Abelian* 2*-avoidable.*

For the rest of this paper, we will work exclusively in the Abelian sense. For brevity we therefore usually write 'w avoids p' for 'w avoids p in the Abelian sense' and 'p is k-avoidable' for 'p is Abelian k-avoidable', etc.

We will use the usual notions of combinatorics on words, such as *word, letter, factor (subword), prefix, length, morphism* etc. A standard reference is [10]. For a word w and a letter a, $|w|_a$ denotes the number of occurrences of a in w; for example $|0111|_0 = 1$, $|0111|_1 = 3$.

Suppose that S and T are alphabets, and $\mu : S \to T$ is a bijection. Then μ extends to a morphism from S^* to T^*. We see that a word $q \in S^*$ is avoidable or k-avoidable if and only if μq is avoidable. We say that $\mu(q)$ and q are **isomorphic**.

2. Some morphisms and their fixed points

For each non-negative integer n, we define a morphism $f_n : \{0, 1\}^* \to \{0, 1\}^*$ generated by

$$f_n(0) = 0^{n+1}1, \qquad f_n(1) = 01^n.$$

When $n = 2$, we have Dekking's [6] morphism $f_2(0) = 0001$, $f_2(1) = 011$. Each morphism f_n has a fixed point w_n, *viz.*, an infinite binary word such that $w_n = f_n(w_n)$. Since $f_n(0), f_n(1)$ both start with 0, this unique fixed point is $w_n = \lim_{m\to\infty} f_n^m(0)$. In [5, Theorem 3.7] we showed that for $n \geq 2$, w_n avoids x^{n+2} in the Abelian sense. Presently we are interested in w_2, w_3.

Let $D_n = n^2 + n - 1$. For $n \geq 2$, we define a weight function $g_n : \{0, 1\}^* \to \mathbb{Z}_{D_n}$ generated by

$$g_n(0) = n, \qquad g(1) = -1.$$

Under g_n, the proper prefixes of $f_n(0)$ map to $n, 2n, 3n, \ldots, n^2, n^2 + n$ ($\equiv 1 \bmod D_n$). On the other hand, the proper prefixes of $f_n(1)$ have weights $n, n-1, n-2, \ldots, 1$. Let $\Pi_n = \{0, 1, 2, \ldots, n\} \cup \{2n, 3n, \ldots, n^2\}$, the set of weights of prefixes of $f_n(0), f_n(1)$.

Combining Corollary 2.8, Corollary 3.11 and Lemma 3.12 in [5], gives the following lemma:

Lemma 2.1. *Let $n \geq 2$ be an integer. Let $q = q_1 q_2 \ldots q_m$, $q_i \in \{0, 1\}$, $1 \leq i \leq m$. Suppose that $|q|_0, |q|_1 \geq n+2$. If w_n contains an Abelian instance of q, then there is an integer $n_0 \in \Pi_n$ and a weight function $h : \{0, 1\} \to \mathbb{Z}_{D_n}$, where $h(0), h(1) \neq 0$, $h(0) \neq h(1)$, such that*

$$n_0 + \sum_{j=1}^{s} h(q_j) \in \Pi_n, \quad 1 \leq s \leq m. \tag{1}$$

3. Unavoidable binary patterns as walks on graphs

Fix $n \geq 2$ and suppose that $a, b \in \mathbb{Z}_{D_n} - \{0\}$. Consider the directed graph $G(n, \{a, b\})$ with vertex set Π_n, and

- an edge from u to v labelled by a whenever $u + a \equiv v \pmod{D_n}$
- an edge from u to v labelled by b whenever $u + b \equiv v \pmod{D_n}$.

The case when $n = 3$, $\{a, b\} = \{2, 4\}$ is illustrated in Fig. 1.

Let $q = q_1 q_2 \ldots q_m \in \{0, 1\}^*$ be a pattern such that w_n contains an Abelian instance of q. If $h(0)$, $h(1)$ and n_0 are as in Lemma 2.1, then the sequence $n_0, n_0 + h(q_1), n_0 + h(q_1) + h(q_2), \ldots, n_0 + \sum_{j=1}^{m} h(q_j)$ is a path in $G(n, \{h(0), h(1)\})$, and the sequence of edge labels on this path is given by the word $h(q) = h(q_1)h(q_2)\cdots h(q_m)$ over alphabet $\{h(a), h(b)\}$. We see

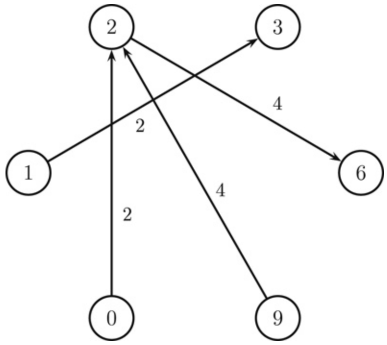


Fig. 1. The graph $G(3, \{2, 4\})$.

that $h(q)$ is isomorphic to q. If we relabel the edges of $G(n, \{a, b\})$ using the substitution $h(0) \to 0, h(1) \to 1$, then q can be walked on $G(n, \{a, b\})$.

Definition 3.1. Let $q \in \{0, 1\}^*$. Let G be a directed graph with edge labels from $\{a, b\}$. We say that q **can be walked on** G if there is a pattern $p \in \{a, b\}^*$ isomorphic to q, such that p labels a walk on G.

Clearly, the precise nature of the labels a and b in this definition is irrelevant, and we may replace a and b by $\mu(a), \mu(b)$ where μ is any bijection.

We see that a longest path in $G(3, \{2, 4\})$ has length 2 (either $0 \to 2 \to 6$ or $9 \to 2 \to 6$). It follows that if $\{h(0), h(1)\} = \{2, 4\}$, then $|q| \leq 2$. If we carefully use observations of this sort, the graphs $G(3, \{a, b\})$ give us a tool for showing that long binary words are avoided either by w_2 or by w_3. We establish the following lemma:

Lemma 3.2. *Let $q \in \{0, 1\}^*$. At least one of the following holds:*

1. *The word q encounters xxxx, so that q is avoided by w_2.*
2. *There are no values $a, b \in \mathbb{Z}_{11} - \{0\}$ with $a \neq b$ such that word q can be walked on $G(3, \{a, b\})$. In this case q is avoided by w_3.*
3. *We have $|q| < 118$.*

Our main result (Theorem 1.2), which states that binary patterns of length greater than 118 are Abelian 2-avoidable, follows as a corollary to Lemma 3.2.

4. Proof of Lemma 3.2

No infinite walks (in fact, no walks of length greater than 2) are possible in the graph $G(3, \{2, 4\})$ in Fig. 1. Several of the $G(3, \{a, b\})$ have this property. We make use of the following fact:

Observation 4.1. *Let G be a directed graph with vertices $v_1, v_2, \ldots, v_r$. Let A_G be an adjacency matrix of G; that is, $A_G = [a_{ij}]_{r\times r}$,*

$$a_{ij} = \begin{cases} 1, & G \text{ contains an edge from } v_i \text{ to } v_j \\ 0, & \text{otherwise.} \end{cases}$$

So if $(A_G)^s = [0]_{r\times r}$ for some positive integer s, then there are no walks of length s on G.

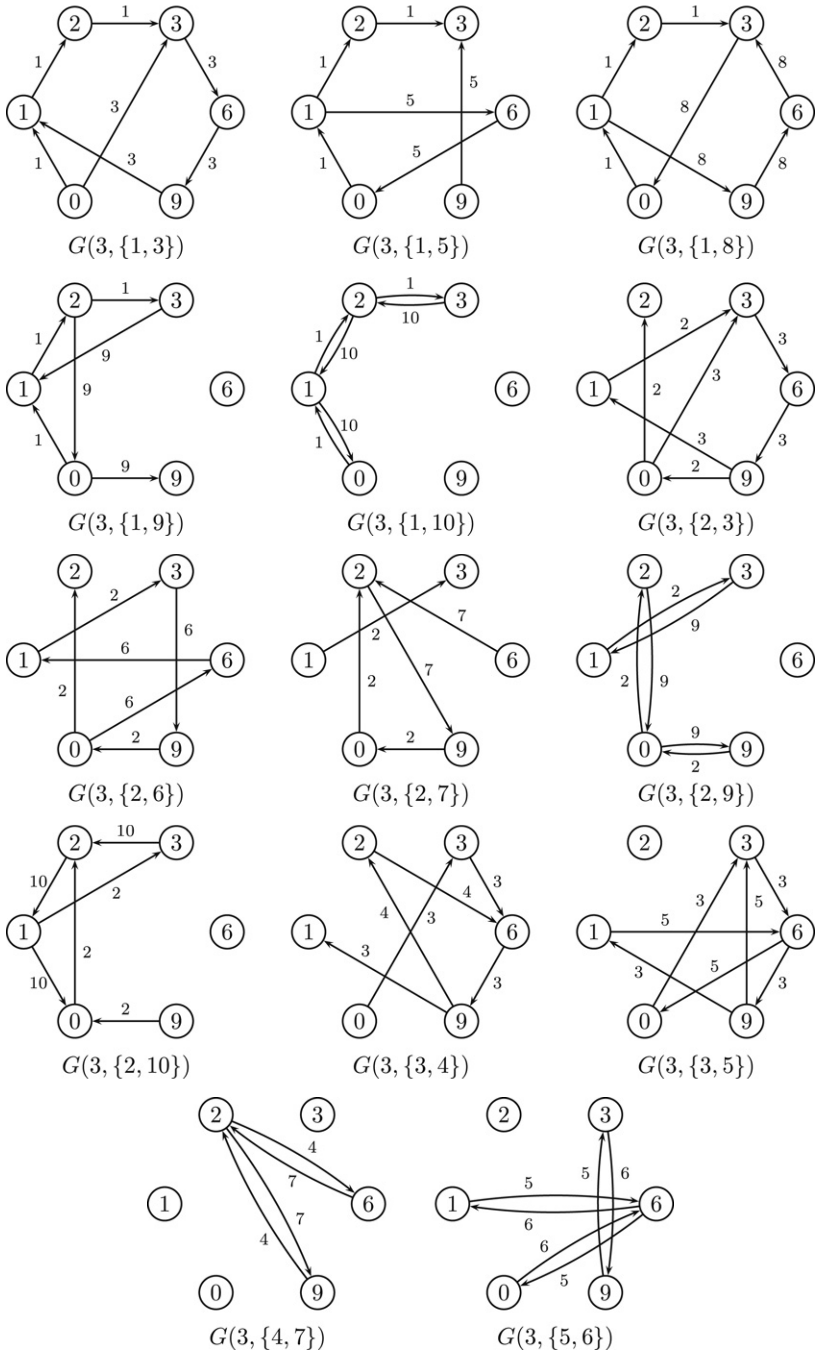
Using this observation with $s = 6$, one computes for which pairs $\{a, b\}, a, b \in \mathbb{Z}_{11} - \{0\}, a < b$, the graph $G(3, \{a, b\})$ allows walks of length 6 or more. This set of pairs is

$$\begin{aligned}\mathcal{P} = \big\{&\{1, 3\}, \{1, 5\}, \{1, 8\}, \{1, 9\}, \{1, 10\}, \{2, 3\}, \{2, 6\}, \{2, 7\}, \{2, 9\},\\ &\{2, 10\}, \{3, 4\}, \{3, 5\}, \{3, 8\}, \{3, 10\}, \{4, 7\}, \{4, 9\}, \{5, 6\}, \{5, 9\},\\ &\{6, 8\}, \{6, 10\}, \{7, 8\}, \{8, 9\}, \{8, 10\}\big\}.\end{aligned}$$

One notices the symmetry $\{a, b\} \in \mathcal{P} \Leftrightarrow \{11 - a, 11 - b\} \in \mathcal{P}$, since walks in $G(3, \{a, b\})$ correspond to the reverses of walks in $G(3, \{11 - a, 11 - b\})$ in the obvious way. Reducing $\mathcal{P}$ with respect to this symmetry leaves the pairs

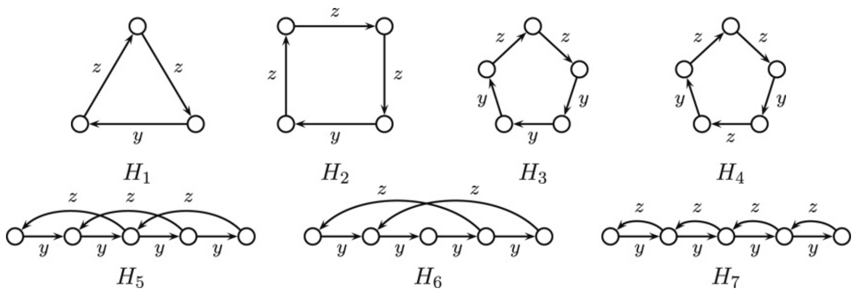
$$\begin{aligned}\mathcal{Q} = \big\{&\{1, 3\}, \{1, 5\}, \{1, 8\}, \{1, 9\}, \{1, 10\}, \{2, 3\}, \{2, 6\}, \{2, 7\}, \{2, 9\},\\ &\{3, 4\}, \{3, 5\}, \{3, 8\}, \{4, 7\}, \{5, 6\}\big\}.\end{aligned}$$

If w_3 encounters pattern q with $|q| \geq 6$, then q may be walked on $G(3, \{a, b\})$, for some $\{a, b\} \in \mathcal{Q}$. The graphs are illustrated below:



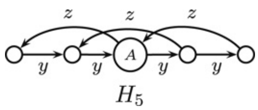
Consider the graph $G(3, \{1, 5\})$. The only non-empty walk w visiting vertex 9 is the walk of length 1, $w = 5$. Similarly, any walk on this graph visits each of vertex 2 and vertex 3 at most once. Thus, if q can be walked on $G(3, \{1, 5\})$, and $|q| > 2$, then q has a factor q' with $|q'| \geq |q| - 2$ such that q' can be walked on the directed triangle on vertices 1, 6 and 0 with the adjacencies $1 \to 6 \to 0 \to 1$. Thus q' is a factor of $(551)^{\omega}$.

Similar analysis of the other $G(3, \{a, b\})$ graphs (as illustrated) shows that if w_3 encounters pattern q with $|q| \geq 6$, then q has a factor q' with $|q'| \geq |q| - 2$ where q' can be walked on one of the following graphs, given appropriate choices of y and z:



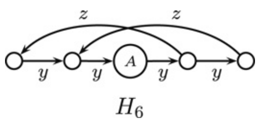
Since H_1 is a directed cycle, any walk q' of length at least 12 on H_1 contains an Abelian instance of *xxxx*. Thus q' is avoided by w_2 since w_2 avoids *xxxx*. Similarly any walk of length at least 20 on H_2, H_3 or H_4 is avoided by w_2.

A slightly longer argument deals with walks on H_5. Label the central vertex A:



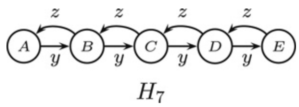
If vertex A is deleted, H_5 allows only walks of length 1. However, if v labels a walk in H_5 from A to A, then $u \sim yyz$. Thus any walk q' on H_5 of length at least 14 contains a factor $v_1v_2v_3v_4$ with each $v_i \sim yyz$, so that q' contains an Abelian instance of *xxxx*.

Similarly, consider H_6 with the central vertex labelled A:



If vertex A is deleted, H_6 allows only walks of length 2. However, if v labels a walk in H_6 from A to A, then $v \sim yyyz$. Thus any walk q' on H_6 of length at least 20 contains a factor $v_1v_2v_3v_4$ with each $v_i \sim yyyz$, so that q' contains an Abelian instance of *xxxx*.

Now we deal with H_7. Label the vertices of H_7 by A, B, C, D and E as below. Consider an infinite walk on H_7 containing no Abelian instance of *xxxx*. Rather than focusing on edge labels, consider the sequence of vertex labels given by this walk, $w = w_1w_2w_3\cdots$ where each $w_i \in \{A, B, C, D, E\}$. Let $u = u_1u_2u_3\cdots$ be the sequence of edge labels associated with w, that is u_i labels the edge from w_i to w_{i+1}. Considering H_7, we see that if $w_i = w_j$ for some $j > i$, then $|u_iu_2\cdots u_{j-1}|_y = |u_iu_2\cdots u_{j-1}|_z$.



By van der Waerden's Theorem, there are positive integers n and m such that

$$w_n = w_{n+m} = w_{n+2m} = w_{n+3m} = w_{n+4m}.$$

This implies that

$$\begin{aligned}|u_{n+(i-1)m}u_{n+(i-1)m+1}\cdots u_{n+im-1}|_y &= m/2\\ &= |u_{n+(i-1)m}u_{n+(i-1)m+1}\cdots u_{n+im-1}|_z\end{aligned}$$

for $i = 1, 2, 3, 4$, so that

$$u_{n+(i-1)m}u_{n+(i-1)m+1}\cdots u_{n+im-1} \sim u_{n+im}u_{n+im+1}\cdots u_{n+(i+1)m-1}$$

for $i = 1, 2, 3$, and v contains an Abelian instance of *xxxx*, which is a contradiction. It follows that there is a positive integer N_0 such that if q' labels edges of H_7, and q' contains no Abelian instance of *xxxx*, then $|q'| < N_0$.

This analytic argument would establish Lemma 3.2, but with 118 replaced by $N = 2 + \max(20, N_0)$. Computer search establishes that the longest walks on H_7 not containing Abelian instances of *xxxx* have length 118. There are 6 such walks, of which one is

xyxyxxxyxyyxxxyxyxyyyxyyxxyxxyyxxxyyxyy
xxyyyxxxyyyxyxyxxxyxyyxxxyxyxyyyxxxyyyx
xyyxyyxxxyyxxyxxyyxyyyxyxyxxxyxyyxxxyxyx.

This completes the proof of Lemma 3.2.

5. Open problems

The following problems are in ascending order of presumed difficulty:

1. Characterise which binary patterns are avoided in the Abelian sense by w_2.
2. Characterise which binary patterns are 2-avoidable in the Abelian sense.
3. For each k, characterise which binary patterns are k-avoidable in the Abelian sense.

Acknowledgement

The authors are supported by NSERC Discovery Grants.

References

[1] D.R. Bean, A. Ehrenfeucht and G. McNulty, Avoidable patterns in strings of symbols, Pacific J. Math. 85 (1979) 261–294.
[2] J. Berstel, Axel Thue's papers on repetitions in words: A translation, Publications du LaCIM, 20, Université du Quebec à Montréal, 1994.
[3] J. Cassaigne, Motifs évitables et régularités dans les mots, thèse de doctorat, L.I.T.P., Université Paris 6, 1994.
[4] J.D. Currie, V. Linek, Avoiding patterns in the Abelian sense, Canad. J. Math. 53 (2001) 696–714.
[5] J.D. Currie, T.I. Visentin, On Abelian 2-avoidable binary patterns, Acta Inform. 43 (2007) 521–533.
[6] F.M. Dekking, Strongly non-repetitive sequences and progression-free sets, J. Combin. Theory Ser. A 27 (1979) 181–185.
[7] P. Erdös, Some unsolved problems, Magyar Tud. Akad. Mat. Kutato. Int. Kozl. 6 (1961) 221–254.
[8] A.A. Evdomikov, Strongly asymmetric sequences generated by a finite number of symbols, Dokl. Akad. Nauk. SSSR 179 (1968) 1268–1271; Soviet Math. Dokl. 9 (1968) 536–539.
[9] V. Keränen, Abelian squares are avoidable on 4 letters, in: Automata, Languages and Programming: Lecture notes in Computer Science, vol. 623, Springer-Verlag, 1992, pp. 41–52.
[10] M. Lothaire, Combinatorics on Words, in: Encyclopedia of Mathematics and its Applications, vol. 17, Addison-Wesley, Reading, Massachusetts, 1983.
[11] P.A.B. Pleasants, Non-repetitive sequences, Proc. Cambridge Philos. Soc. 68 (1970) 267–274.
[12] P. Roth, Every binary pattern of length six is avoidable on the two-letter alphabet, Acta Inform. 29 (1992) 95–107.
[13] A. Thue, Über unendliche Zeichenreihen, Norske Vid. Selsk. Skr. I. Mat. Nat. Kl. Christiana (7) (1906).
[14] A. Zimin, Blocking sets of terms, Mat. Sb. (N.S.) 119 (161) (1982); Math. USSR Sbornik 47 (1984) 353–364.