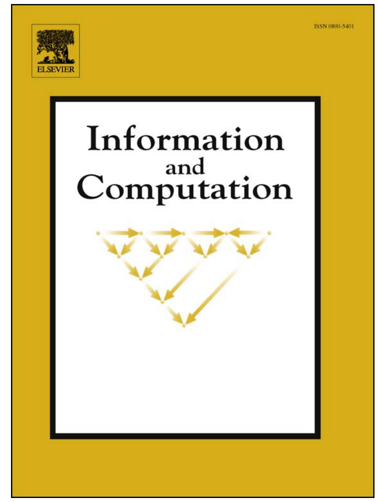
Aistis Atminas, Vadim Lozin, Mikhail Moshkov

WQO is decidable for factorial languages

Accepted Manuscript

WQO is Decidable for Factorial Languages

Aistis Atminas, Vadim Lozin, Mikhail Moshkov



PII: S0890-5401(17)30139-6
DOI: http://dx.doi.org/10.1016/j.ic.2017.08.001
Reference: YINCO 4300

To appear in: *Information and Computation*

Received date: 28 March 2015
Revised date: 29 May 2017

Please cite this article in press as: A. Atminas et al., WQO is Decidable for Factorial Languages, *Inf. Comput.* (2017), http://dx.doi.org/10.1016/j.ic.2017.08.001

WQO is Decidable for Factorial Languages*

Aistis Atminas† Vadim Lozin‡ Mikhail Moshkov§

Abstract

A language is factorial if it is closed under taking factors, i.e. contiguous subwords. Every factorial language can be described by an antidictionary, i.e. a minimal set of forbidden factors. We show that the problem of deciding whether a factorial language given by a finite antidictionary is well-quasi-ordered under the factor containment relation can be solved in polynomial time. We also discuss possible ways to extend our solution to permutations and graphs.

Keywords: well-quasi-ordering; factorial language; polynomial-time algorithm; induced subgraph; permutation

1 Introduction

Well-quasi-ordering (wqo) is a highly desirable property and frequently discovered concept in mathematics and theoretical computer science [10, 13]. One of the most remarkable recent results in this area is the proof of Wagner's conjecture stating that the set of all finite graphs is well-quasi-ordered by the minor relation [16]. However, the subgraph or induced subgraph relation is not a well-quasi-order. Other examples of important relations that are not well-quasi-orders are pattern containment relation on permutations [17], embeddability relation on tournaments [5], minor ordering of matroids [11], factor (contiguous subword) relation on words [8]. On the other hand, each of these relations may become a well-quasi-order under some additional restrictions. In the present paper, we study restrictions given in the form of obstructions, i.e. minimal excluded ("forbidden") elements (precise definitions and examples will be given in the next section). The fundamental problem of our interest is the following: given a partial order P and a *finite* set Z of obstructions, determine if the set of elements of P containing no elements from Z forms a well-quasi-order. This problem was studied for the induced subgraph relation on graphs [12], the pattern containment relation on permutations [4], the embeddability relation on tournaments [5], the minor ordering of matroids [11]. However, the decidability of this problem has been shown only for one or two forbidden elements (graphs, permutations, tournaments,

*An extended abstract of this paper appeared in the proceedings of LATA 2013 [3].

†Department of Mathematics, London School of Economics, Houghton Street, London WC2A 2AE, UK. Email: A.Atminas@lse.ac.uk.

‡Mathematics Institute, University of Warwick, Coventry, CV4 7AL, UK. Email: V.Lozin@warwick.ac.uk.
This author gratefully acknowledges support from EPSRC, grant EP/L020408/1.

§CEMSE Division, King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Saudi Arabia. Email: mikhail.moshkov@kaust.edu.sa. Support of KAUST is gratefully acknowledged.

matroids). Whether this problem is decidable for larger numbers of forbidden elements is an open question. In the present paper, we answer this question positively for factorial languages.

A language is factorial if it is closed under taking factors. Every factorial language can be described by an antidictionary, i.e. a minimal set of forbidden factors. The main result of the paper, presented in Sections 3 and 4, is the proof of polynomial-time solvability of the problem of deciding whether a factorial language given by a finite antidictionary is well-quasi-ordered under the factor containment relation. We also discuss possible ways to extend our solution to permutations and graphs in Sections 5, 6, 7. All preliminary information related to the topic of the paper can be found in Section 2.

2 Preliminaries

2.1 Partial orders and WQO

For a set A we denote by A^2 the set of all ordered pairs of (not necessarily distinct) elements from A. A *binary relation* on A is a subset of A^2. If a binary relation $\mathcal{R} \subset A^2$ is

- *reflexive*, i.e. $(a,a) \in \mathcal{R}$ for each $a \in A$,
- *transitive*, i.e. $(a,b) \in \mathcal{R}$ and $(b,c) \in \mathcal{R}$ imply $(a,c) \in \mathcal{R}$,

then $\mathcal{R}$ is a *quasi-order* (also known as *pre-order*). If additionally $\mathcal{R}$ is

- *antisymmetric*, i.e. $(a,b) \in \mathcal{R}$ and $(b,a) \in \mathcal{R}$ imply $a = b$,

then $\mathcal{R}$ is a *partial order*.

We say that two elements $a, b \in A$ are comparable with respect to $\mathcal{R}$ if either $(a,b) \in \mathcal{R}$ or $(b,a) \in \mathcal{R}$. A set of pairwise comparable elements of A is called a *chain* and a set of pairwise incomparable elements of A is called an *antichain*.

A quasi-ordered set is *well-quasi-ordered* if it contains

- neither infinite strictly decreasing chains, in which case we say that the set is *well-founded*,
- nor infinite antichains.

All examples of quasi-orders in this paper will be antisymmetric (i.e. partial orders) and well-founded, in which case well-quasi-orderability is equivalent to the non-existence of infinite antichains.

Examples.

(1) Let A be the set of all finite simple (i.e. undirected, without loops and multiple edges) graphs. If a graph $H \in A$ can be obtained from a graph $G \in A$ by a (possibly empty) sequence of

  – vertex deletions, then H is an induced subgraph of G,
  – vertex deletions and edge deletions, then H is a subgraph of G,

– vertex deletions, edge deletions and edge contractions,[1] then H is a minor of G,
– vertex deletions and edge contractions, then H is an induced minor of G.

According to the result of Robertson and Seymour [16] the minor relation on the set of graphs is a well-quasi-order. However, this is not the case for the subgraph, induced subgraph and induced minor relation. Indeed, it is not difficult to see that the set of all chordless cycles $C_3, C_4, C_5, \ldots$ creates an infinite antichain with respect to both subgraph and induced subgraph relations, and the complements of the cycles form an infinite antichain with respect to the induced minor relation. Besides, the set of so-called H-graphs (i.e. graphs represented in Figure 1) also forms an infinite antichain with respect to subgraph and induced subgraph relations.

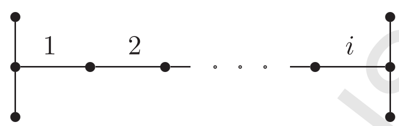


Figure 1: The graph H_i

(2) Let A be the set of all finite permutations. We say that a permutation $\pi \in A$ of n elements is contained in a permutation $\rho \in A$ of k elements ($n \leq k$) as a *pattern*, if π can be obtained from ρ by deleting some (possibly none) elements and renaming the remaining elements consecutively in the increasing order. Obviously, the pattern containment relation is a well-founded partial order. However, whether it is a well-quasi-order is not an obvious fact. Finding an infinite antichain of permutations becomes much easier if we associate to each permutation its permutation graph. Let π be a permutation on the set $N = \{1, 2, \ldots, n\}$. The permutation graph G_π of π is the graph with vertex set N in which two vertices i and j are adjacent if and only if they form an inversion in π (i.e. $i < j$ and $\pi(i) > \pi(j)$). It is not difficult to see that if ρ contains π as a pattern, then G_ρ contains G_π as an induced subgraph. Therefore, if $G_1, G_2, \ldots$ is an infinite antichain of permutation graphs with respect to the induced subgraph relation, then the corresponding permutations form an infinite antichain with respect to the pattern containment relation. Since the H-graphs (Figure 1) are permutation graphs (which is easy to see), we conclude that the pattern containment relation on permutations is not a well-quasi-order.

(3) Let A be the set of all finite words in a finite alphabet. A word $\alpha \in A$ is said to be a *factor* of a word $\beta \in A$ if α can be obtained from β by omitting a (possibly empty) suffix and prefix. If the alphabet contains at least two symbols, say 1 and 0, the factor containment relation (also known as the infix order) is not a well-quasi-order, since it necessarily contains an infinite antichain, for instance, 010, 0110, 01110, etc.

[1] Edge contraction is the operation of replacing two adjacent vertices a and b by a new vertex, which is adjacent to every neighbour of a and to every neighbour of b in the rest of the graph

2.2 Hereditary properties of partial orders

Let $(A, \mathcal{R})$ be a well-founded partial order. A property on A is a subset of A. A property $P \subseteq A$ is *hereditary* (with respect to $\mathcal{R}$) if $x \in P$ implies $y \in P$ for every $y \in A$ such that $(y, x) \in \mathcal{R}$. Hereditary properties are also known as lower ideals or downward closed sets.

Examples.

- If A is the set of all finite graphs and $\mathcal{R}$ is the minor relation, then a hereditary property on A is known as a *minor-closed* class of graphs.
- If A is the set of all finite graphs and $\mathcal{R}$ is the subgraph relation, then a hereditary property on A is known as a *monotone* class of graphs.
- If A is the set of all permutations and $\mathcal{R}$ is the pattern containment relation, then a hereditary property on A is known as a *pattern* class or *pattern avoiding* class.
- If A is the set of all words in a finite alphabet and $\mathcal{R}$ is the factor containment relation, then a hereditary property on A is known as a *factorial language*.

The word "avoiding" used in the terminology of permutations suggests that a hereditary property can be described in terms of "forbidden" elements. To better explain this idea, let us introduce the following notation: given a set $Z \subseteq A$, we denote

$$Free(Z) := \{a \in A \mid (z, a) \notin \mathcal{R} \ \forall z \in Z\}.$$

Obviously, for any $Z \subseteq A$, the set $Free(Z)$ is hereditary. On the other hand, for any hereditary property $P \subseteq A$ there is a unique minimal set $Z \subseteq A$ such that $P = Free(Z)$. We call Z the set of *forbidden elements* for $Free(Z)$ and observe that a minimal set of forbidden elements is necessarily an antichain.

Examples.

- Since the minor relation on graphs contains no infinite antichains, any minor-closed class of graphs can be described by a *finite* set of forbidden minors. In particular, for the class of planar graphs the set of minimal forbidden minors consists of K_5, the complete graph on 5 vertices, and $K_{3,3}$, the complete bipartite graph with 3 vertices in each part.
- The set of minimal forbidden permutations for a pattern avoiding class is also known as the *base* of the class.
- The set of minimal forbidden words for a factorial language is also known as the *antidictionary* of the language.

In the above notation, the problem of our interest can be stated as follows:

Problem 1. *Given a finite set $Z \subset A$, determine if $Free(Z)$ is well-quasi-ordered with respect to $\mathcal{R}$.*

This question is not applicable to the minor relation on graphs, since this relation is a well-quasi-order. For hereditary properties of graphs with respect to the subgraph relation, Problem 1 has a simple solution which is due to Ding [9]: a monotone class of graphs is well-quasi-ordered by the subgraph relation if and only if it contains finitely many cycles and finitely many H-graphs. Therefore, if Z is finite, then $Free(Z)$ is well-quasi-ordered with respect to the subgraph relation if and only if Z includes a chordless path (or an induced subgraph of a chordless path), because otherwise $Free(Z)$ contains infinitely many cycles.

For other relations, such as the induced subgraph relation on graphs or pattern containment relation on permutations, only partial results are available, where Z contains one or two elements (see e.g. [1, 12]). Whether this problem is decidable for larger numbers of forbidden elements is an open question. In the present paper, we study Problem 1 for factorial languages and show that the problem is efficiently solvable for any finite set Z. To this end, we use the result from [6] which allows representing a factorial language defined by a finite antidictionary in the form of a deterministic finite automaton.

2.3 Languages and automata

Let $k \geq 2$ be a natural number and $E_k = \{0, 1, \ldots, k-1\}$ be an alphabet. A *deterministic finite automaton* over E_k is a triple $\mathcal{A} = (G, q_0, Q)$, where

- G is a finite directed graph, possibly with multiple edges and loops, in which the edges are labeled with letters from E_k in such a way that any two edges leaving the same node have different labels,
- q_0 is a node of G, called the *start node*, and
- Q is a nonempty set of nodes of G, called the *terminal nodes*.

A directed path in G is any sequence $v_1, e_1, \ldots, v_m, e_m, v_{m+1}$ of nodes v_i and edges e_j such that for each $j = 1, \ldots, m$, the edge e_j is directed from v_j to v_{j+1}. We emphasize that both nodes and edges can appear in such a path repeatedly.

With each directed path τ in the graph G we associate a word over E_k by reading the labels of the edges of τ (listed along the path) and denote this word by $w(\tau)$. A directed path in G will be called an $\mathcal{A}$-path if it starts at the node q_0 and ends at a terminal node.

Let α be a word over E_k. We say that an automaton $\mathcal{A} = (G, q_0, Q)$ accepts α if there is an $\mathcal{A}$-path τ such that $w(\tau) = \alpha$. The set of all words accepted by $\mathcal{A}$ is called the language accepted (or recognized) by $\mathcal{A}$ and this language is denoted $L(\mathcal{A})$. It is well-known that the languages accepted by deterministic finite automata are precisely the regular languages.

The following theorem is a corollary of Proposition 5 in [6].

Theorem 1. *Given a set $Z = \{w_1, \ldots, w_n\}$ of pairwise incomparable words over E_k, in time polynomial in $|w_1| + \ldots + |w_n|$ and k one can construct a deterministic finite automaton $\mathcal{A}$ such that $L(\mathcal{A})$ coincides with the factorial language $Free(Z)$.*

We call an automaton $\mathcal{A} = (G, q_0, Q)$ *reduced* if for each node of G there exists an $\mathcal{A}$-path containing this node. It is not difficult to see that any deterministic finite automaton can be transformed into an equivalent (i.e. accepting the same language) reduced deterministic finite automaton in polynomial time. This observation together with Theorem 1 reduce Problem 1 to the following one:

Problem 2. *Given a reduced deterministic finite automaton $\mathcal{A}$, determine if $L(\mathcal{A})$ is well-quasi-ordered with respect to the factor containment relation.*

In the next two sections, we give an efficient solution to this problem.

3 Auxiliary results

Given a word α, we denote by $|\alpha|$ the length of α, i.e. the number of letters in α. Also, α^i denotes concatenation of i copies of α and is called the i-th power of α.

A word $\alpha = \alpha_1 \dots \alpha_n$ is called a *periodic word with period p* if

- either $p \geq n$
- or $p < n$ and $\alpha_i = \alpha_{i+p}$ for $i = 1, \dots, n-p$.

A word γ will be called a *left extension of a power of a word* α if γ can be represented in the form $\sigma\alpha^i$, where σ is a suffix of α and $i \geq 0$. Similarly, γ will be called a *right extension of a power* of α if γ can be represented in the form $\alpha^i\sigma$, where σ is a prefix of α. Directly from the definition we obtain the following conclusion.

Lemma 1. *A word γ is a left extension of a power of α if and only if the word $\gamma\alpha$ is a periodic word with period $|\alpha|$. A word γ is a right extension of a power of α if and only if the word $\alpha\gamma$ is a periodic word with period $|\alpha|$.*

Now we prove a number of further auxiliary results.

Lemma 2. *Let γ, α, δ be words such that either γ is a left extension of a power of α or δ is a right extension of a power of α. Then the set $\{\gamma\alpha^i\delta : i = 0, 1, 2, \dots\}$ is a chain, i.e. any two words in this set are comparable.*

Proof. Let $\gamma\alpha^i\delta$ and $\gamma\alpha^j\delta$ be two words with $i < j$. If γ is a left extension of a power of α, then the word γ can be represented as $\sigma\alpha^k$ where σ is a suffix of α. Therefore, the word $\gamma\alpha^i\delta$ can be obtained from the word $\gamma\alpha^j\delta$ by removing a prefix of length $(j-i)|\alpha|$. Similarly, if δ is a right extension of a power of α, then the word $\gamma\alpha^i\delta$ can be obtained from the word $\gamma\alpha^j\delta$ by removing a suffix of length $(j-i)|\alpha|$. $\square$

Lemma 3. *Let γ, α, δ be words such that γ is not a left extension of a power of α, and δ is not a right extension of a power of α. Then the set of words $\{\gamma\alpha^i\delta : i = 1, 2, \dots\}$ is an infinite antichain.*

Proof. Let $\alpha = a_1a_2 \dots a_p$ be a word of length p. Suppose there is a contiguous embedding $\phi : \gamma\alpha^j\delta \to \gamma\alpha^k\delta$. Then the factor α^j of the first word is embedded into the factor α^k of the second word. Denote $\alpha^k := a_1a_2 \dots a_{kp}$, where $a_i = a_{i(mod\ p)}$. Suppose the first letter a_1 is mapped by ϕ to a_{h+1}. Then, $a_m = a_{m+h}$ for $m = 1, 2 \dots p$. Now $a_m = a_{m+ih(mod\ p)}$

for $i = 1, 2, \ldots$. We define $p' = gcd(h, p)$, and claim that α is a periodic word of period p'. Indeed, by Bézout's identity we have $p' = xh - yp$ for some positive integers x and y. Hence $a_m = a_{m+xh(mod\ p)} = a_{m+p'}$ for all $m = 1, 2, \ldots, (p - p')$. This shows that α is periodic word with period p'.

Define $\alpha' = a_1 a_2 \ldots a_{p'}$, and write $\alpha = (\alpha')^{p/p'}$. Notice that $h = zp'$ for some nonnegative integer z. If $z \neq 0$, we conclude that the factor γ of the first word is embedded to the suffix of the word $\gamma(\alpha')^z$ taken from the second word. Adding extra powers of $(\alpha')^z$ to both words we conclude that $\gamma(\alpha')^z$ is a suffix of $\gamma(\alpha')^{2z}$, $\gamma(\alpha')^{2z}$ is a suffix of $\gamma(\alpha')^{3z}$, $\ldots$, $\gamma(\alpha')^{(s-1)z}$ is a suffix of $\gamma(\alpha')^{sz}$. So we conclude that γ is a suffix of $\gamma(\alpha')^z$ which is a suffix of $\gamma(\alpha')^{2z}$, which is a suffix of $\gamma(\alpha')^{3z}$, ..., which is a suffix of $\gamma(\alpha')^{sz}$. So γ is a suffix of $\gamma(\alpha')^{sz}$. For s larger than $\frac{|\gamma|}{|(\alpha')^z|}$ we conclude that γ is a suffix of $(\alpha')^{sz}$ which means that it is a left extension of a power of α' and hence a left extension of a power of α. This contradiction shows that $z = 0$, and hence $h = 0$, and that the first word is a prefix of the second one.

So consider the case when the first word $\gamma\alpha^j\delta$ is the prefix of the second $\gamma\alpha^k\delta$. Then, in particular, δ from the first word is mapped to the prefix of $\alpha^{(k-j)}\delta$ from the second word. Now, adding the powers of $\alpha^{(k-j)}$, we conclude that $\alpha^{(k-j)}\delta$ is a prefix of $\alpha^{2(k-j)}\delta$, $\alpha^{2(k-j)}\delta$ is a prefix of $\alpha^{3(k-j)}\delta$, $\ldots$, $\alpha^{(s-1)(k-j)}\delta$ is a prefix of $\alpha^{s(k-j)}\delta$. So δ is a prefix of $(\alpha)^{(k-j)s}\delta$. If $k - j > 0$ then we choose s to be larger than $\frac{|\delta|}{|(\alpha)^{(k-j)}|}$ and we conclude that δ is a prefix of $\alpha^{(j-k)s}$. But this means that δ is a right extension of a power of α. This contradiction shows that $k = j$ and hence no two different words are comparable in the set $\{\gamma\alpha^i\delta : i = 1, 2, \ldots\}$. Therefore, this set is an antichain. □

Lemma 4. *Let $\gamma, \alpha, \delta, \beta$ be words. If δ is not a right extension of a power of α, then the word $\gamma\alpha^{|\beta|}\delta$ is not a left extension of a power of β.*

Proof. Let us assume the contrary, i.e. assume that $\gamma\alpha^{|\beta|}\delta$ is a left extension of a power of β. Then, by Lemma 1, the word $\mu = \gamma\alpha^{|\beta|}\delta\beta$ is a periodic word with period $|\beta|$ and, therefore, with period $|\beta|\,|\alpha|$. Since $\alpha^{|\beta|}$ is a factor of μ and $|\beta|\,|\alpha|$ is a period of μ, the word μ is a factor of the word $\alpha^{|\beta|p}$ for a large enough natural number p. Hence, μ is a periodic word with period $|\alpha|$. As a result, $\alpha\delta$ also is a periodic word with period $|\alpha|$. However, this is impossible by Lemma 1, since δ is not a right extension of a power of α. □

Lemma 5. *Let $\alpha_1, \alpha_2, \gamma, \sigma, \delta$ be words. If σ is a suffix of α_1 and δ is a prefix of α_2, then the set $S = \{\sigma\alpha_1^i\gamma\alpha_2^j\delta \mid i, j \in \mathbb{N}\}$ is well-quasi-ordered by the factor containment relation.*

Proof. Let $\mu_1 = \sigma\alpha_1^{i_1}\gamma\alpha_2^{j_1}\delta$ and $\mu_2 = \sigma\alpha_1^{i_2}\gamma\alpha_2^{j_2}\delta$ be two words in S. Obviously, if these words are incomparable then either $i_1 < i_2$ and $j_1 > j_2$ or $i_1 > i_2$ and $j_1 < j_2$. This implies that there are at most $i_1 + j_1$ pairwise incomparable words in S incomparable with μ_1. Since for each word μ_1 in S there exist only finitely many pairwise incomparable words which are incomparable with μ_1, the set S does not contain an infinite antichain. □

For any set of words S, we denote by $\langle S\rangle$ the set of all factors of words in S. With this notation we extend the result of the previous lemma as follows.

Lemma 6. *Let $\alpha_1, \alpha_2, \gamma$ be words. Then the set $\langle\{\alpha_1^i\gamma\alpha_2^j \mid i, j \in \mathbb{N}\}\rangle$ is well-quasi-ordered by the factor containment relation.*

Proof. Let $A_{1,p}$, G_p, $A_{2,p}$ be the sets of prefixes and $A_{1,s}$, G_s, $A_{2,s}$ the sets of suffixes of the words α_1, γ, α_2, respectively. Clearly, these sets are finite. Then $\langle\{\alpha_1^i\gamma\alpha_2^j \mid i,j \in \mathbb{N}\}\rangle = \bigcup_{\sigma\in A_{1,s},\delta\in A_{2,p}}\{\sigma\alpha_1^i\gamma\alpha_2^j\delta \mid i,j \in \mathbb{N}\}\bigcup_{\gamma_s\in G_s,\delta\in A_{2,p}}\{\gamma_s\alpha_2^j\delta \mid j \in \mathbb{N}\}\bigcup_{\delta'\in A_{2,s},\delta\in A_{2,p}}\{\delta'\alpha_2^j\delta \mid j \in \mathbb{N}\}\bigcup_{\sigma\in A_{1,s},\gamma_p\in G_p}\{\sigma\alpha_1^i\gamma_p \mid i \in \mathbb{N}\}\bigcup_{\sigma\in A_{1,s},\sigma'\in A_{1,p}}\{\sigma\alpha_1^i\sigma' \mid i \in \mathbb{N}\}$ is a finite union of sets which are well-quasi-ordered by Lemma 5 and Lemma 2.

□

4 Main results

In this section, we show how to decide for a given reduced deterministic finite automaton $\mathcal{A} = (G, q_0, Q)$ whether the language $L(\mathcal{A})$ contains an infinite antichain with respect to the factor containment relation or not. Our solution is based on the analysis of the structure of cycles in G.

A cycle in G is any directed path with at least one edge in which the first and the last nodes coincide. A cycle is *simple* if its nodes are pairwise distinct (except for the first node being equal to the last node). Given a simple cycle C, we denote by $|C|$ the *length* of C, i.e. the number of nodes in C. For a node v of C, we denote by $w(C, v)$ the word of length $|C|$ obtained by reading the labels of the edges of C starting from the node v.

We distinguish between two basic cases: the case where G contains two different simple cycles that have at least one node in common and the case where all simple cycles of G are pairwise node disjoint.

Proposition 1. *Let $\mathcal{A} = (G, q_0, Q)$ be a reduced deterministic finite automaton. If G contains two different simple cycles which have a node in common, then the language $L(\mathcal{A})$ contains an infinite antichain.*

Proof. Let C_1 and C_2 be two different simple cycles with a common node v. Since the cycles are different, we may assume without loss of generality that the node of C_1 following v is different from the node of C_2 following v. As a result, the edges of C_1 and C_2 leaving v are labeled with different letters of the alphabet.

We denote $\alpha = w(C_1, v)$ and $\beta = w(C_2, v)$. The words α and β differ in the first letter according to the above assumption.

Since $\mathcal{A}$ is reduced, there exist a directed path ρ from the start node to v and a directed path π from v to a terminal node. Therefore, every word of the form $w(\rho)\beta^{|C_1|}\alpha^i\beta w(\pi)$ $(i = 1, 2, \ldots)$ belongs to the language $L(\mathcal{A})$. Since the words α and β differ in the first letter, we conclude that β, and hence $\beta w(\pi)$, is not a right extension of a power of α. Let us assume that the word $w(\rho)\beta^{|C_1|}$ is a left extension of a power of α. Then the word $\beta^{|C_1|}$ is a left extension of a power of α. By Lemma 1, the word $\beta^{|C_1|}\alpha$ is a periodic word with period $|\alpha|$, and hence β and α have the same first letter which is impossible. Therefore, $\beta^{|C_1|}$ and $w(\rho)\beta^{|C_1|}$ are not left extensions of a power of α and, by Lemma 3, the language $L(\mathcal{A})$ contains an infinite antichain. □

From now on, we consider automata in which every two simple cycles are node disjoint. In this case, we decompose the set of nodes into finitely many subsets of simple structure, called metapaths.

A *metapath* consists of a number of node disjoint simple cycles, say $C_1, \dots, C_t$ (possibly $t = 0$), and a number of directed paths $\rho_0, \dots, \rho_t$ such that ρ_0 connects the start node q_0 to C_1, ρ_1 connects C_1 to C_2, ρ_2 connects C_2 to C_3, and so on, and finally, ρ_t connects C_t to a terminal node of the automaton. Let us observe that ρ_0 and ρ_t can be of length 0, while all the other paths are necessarily of length at least one, since the cycles are node disjoint.

We denote by $s(\rho_i)$ and $f(\rho_i)$ the first and the last node of ρ_i, respectively, and observe that for $i > 0$, $s(\rho_i)$ belongs to C_i, and for $i < t$, $f(\rho_i)$ belongs to C_{i+1}.

If $t = 0$, then the metapath contains no cycles and consists of the path ρ_0 alone. This path connects the start node q_0 to a terminal node of the automaton and no node of this path belongs to a simple cycle.

If $t > 0$, then $f(\rho_0), s(\rho_1), f(\rho_1), \dots, s(\rho_{t-1}), f(\rho_{t-1}), s(\rho_t)$ are the only nodes of the paths $\rho_0, \rho_1, \dots, \rho_t$ that belong to simple cycles.

For $i = 1, \dots, t$, we denote by

- π_i the directed path from $f(\rho_{i-1})$ to $s(\rho_i)$ taken along the cycle C_i,
- γ_i the word $w(\pi_i)w(\rho_i)$,
- α_i the word $w(C_i, f(\rho_{i-1}))$.

Also, by γ_0 we denote the word $w(\rho_0)$.

Let τ be a metapath with t cycles, as defined above. The set $L(\tau)$ of words accepted by this metapath can be described as follows: if $t = 0$ then $L(\tau) = \{\gamma_0\}$, and if $t > 0$ then

$$L(\tau) = \{\gamma_0 \alpha_1^{j_1} \gamma_1 \dots \gamma_{t-1} \alpha_t^{j_t} \gamma_t : j_1, \dots, j_t = 0, 1, \dots\}.$$

Clearly, the set $T(\mathcal{A})$ of all metapaths is finite and

$$L(\mathcal{A}) = \bigcup_{\tau \in T(\mathcal{A})} L(\tau).$$

It is also clear that $L(\mathcal{A})$ contains an infinite antichain if and only if $L(\tau)$ contains an infinite antichain for at least one metapath $\tau \in T(\mathcal{A})$.

If $t = 0$, the set $L(\tau)$ is finite and hence cannot contain an infinite antichain. In order to determine if $L(\tau)$ contains an infinite antichain for $t > 0$, we distinguish between the following three cases: $t = 1$, $t = 2$ and $t \geq 3$. In our analysis below we use the following simple observation:

Observation 1. *For $i = 1, \dots, t-1$, the word γ_i is not a right extension of a power of α_i.*

The validity of this observation is due to the fact that the edge of ρ_i and the edge of C_i leaving vertex $s(\rho_i)$ must have different labels. On the other hand, we note that γ_0 may be a left extension of a power of α_1, while γ_t may be a right extension of a power of α_t.

Observation 2. *The word γ_t is not a right extension of a power of α_t if and only if the length of the path ρ_t is at least 1.*

Proposition 2. *Let τ be a metapath with exactly one cycle. Then $L(\tau)$ contains an infinite antichain if and only if*

- *neither γ_0 is a left extension of a power of α_1*
- *nor γ_1 is a right extension of a power of α_1.*

Proof. If γ_0 is a left extension of a power of α_1 or γ_1 is a right extension of a power of α_1, then $L(\tau)$ does not contain an infinite antichain by Lemma 2.

If neither γ_0 is a left extension of a power of α_1 nor γ_1 is a right extension of a power of α_1, then $L(\tau)$ contains an infinite antichain by Lemma 3. □

Proposition 3. *Let τ be a metapath with exactly two cycles. Then $L(\tau)$ contains an infinite antichain if and only if*

- *either γ_0 is not a left extension of a power of α_1*
- *or γ_2 is not a right extension of a power of α_2.*

Proof. Assume γ_0 is not a left extension of a power of α_1. From Observation 1 we know that γ_1 is not a right extension of a power of α_1. This implies that $\gamma_1\gamma_2$ is not a right extension of a power of α_1. Therefore, by Lemma 3, the set $\{\gamma_0\alpha_1^i\gamma_1\gamma_2 : i = 1, 2, \ldots\}$ is an infinite antichain. Since this set is a subset of $L(\tau)$, we conclude that $L(\tau)$ contains an infinite antichain.

Suppose now that γ_2 is not a right extension of a power of α_2. By Observation 1, γ_1 is not a right extension of a power of α_1. Therefore, by Lemma 4, $\gamma_0\alpha_1^{|\alpha_2|}\gamma_1$ is not a left extension of a power of α_2. This implies, by Lemma 3, that the set $\{\gamma_0\alpha_1^{|\alpha_2|}\gamma_1\alpha_2^i\gamma_2 : i = 1, 2, \ldots\}$ is an infinite antichain. Since this set is a subset of $L(\tau)$, we conclude that $L(\tau)$ contains an infinite antichain.

Finally, assume that γ_0 is a left extension of a power of α_1 and γ_2 is a right extension of a power of α_2. Then the set $L(\tau) = \{\gamma_0\alpha_1^i\gamma_1\alpha_2^j\gamma_2 \mid i, j \in \mathbb{N}\}$ satisfies the conditions of Lemma 5, and therefore is well-quasi-ordered. □

Proposition 4. *Let τ be a metapath with $t \geq 3$ cycles. Then $L(\tau)$ contains an infinite antichain.*

Proof. By Observation 1, γ_1 is not a right extension of a power of α_1, and hence, by Lemma 4, $\gamma_0\alpha_1^{|\alpha_2|}\gamma_1$ is not a left extension of a power of α_2. Also, by Observation 1, γ_2 is not a right extension of a power of α_2, and hence $\gamma_2 \ldots \gamma_t$ is not a right extension of a power of α_2. Therefore, by Lemma 3, the set $\{\gamma_0\alpha_1^{|\alpha_2|}\gamma_1\alpha_2^i\gamma_2 \ldots \gamma_t : i = 0, 1, 2, \ldots\}$ is an infinite antichain. Since this set is a subset of $L(\tau)$, we conclude that $L(\tau)$ contains an infinite antichain. □

We summarize the above discussion in the following final statement which is the main result of the paper.

Theorem 2. *Given a reduced deterministic finite automaton $\mathcal{A} = (G, q_0, Q)$ one can decide in polynomial time whether the language accepted by $\mathcal{A}$ contains an infinite antichain with respect to the factor containment relation or not.*

Proof. Since the automaton $\mathcal{A}$ is finite, the question of the existence of infinite antichains in $L(\mathcal{A})$ is decidable by Propositions 1, 2, 3, 4. Now we prove polynomial-time solvability.

First, we identify strongly connected components in G, which can be done in polynomial time. If at least one strongly connected component contains a simple cycle and is different from the cycle (i.e. contains at least one edge outside of the cycle), then it necessarily contains two cycles with a common node, in which case $L(\mathcal{A})$ contains an infinite antichain by Proposition 1.

If each of the strongly connected components of G is a simple cycle or a single vertex without loops, then the simple cycles of G are pairwise node disjoint. If a node of G belongs to a simple cycle then it will be called a *cyclic* node. Otherwise, it will be called an *acyclic* node.

We now verify if G contains a metapath with at least three cycles. Since $\mathcal{A}$ is reduced, a metapath with at least three cycles exists if and only if there is a directed path in G containing nodes from at least three simple cycles. It is clear that there is a polynomial-time algorithm which, for a given ordered pair of simple cycles C_1 and C_2, verifies if there is a directed path from a node of C_1 to a node of C_2. Using this algorithm, we can check in polynomial time, for each triple of cycles C_1, C_2, C_3, if there is a directed path in G which passes through C_1, C_2, and C_3. If such a path exists then a metapath with at least three cycles exists and, by Proposition 4, $L(\mathcal{A})$ necessarily contains an infinite antichain.

Let us assume now that G does not contain a metapath with three or more cycles. In this case, we should verify if there exists a metapath τ with exactly two cycles satisfying conditions of Proposition 3: $L(\tau) = \{\gamma_0\alpha_1^{j_1}\gamma_1\alpha_2^{j_2}\gamma_2 : j_1, j_2 = 0, 1, \ldots\}$ and either γ_0 is not a left extension of a power of α_1, or γ_2 is not a right extension of a power of α_2. The metapath τ consists of two simple cycles C_1 and C_2, and three directed paths ρ_0, ρ_1, ρ_2 such that ρ_0 connects q_0 to C_1 and has no cyclic nodes with the exception of the last node $f(\rho_0)$, ρ_1 connects C_1 to C_2 and has no cyclic nodes with the exception of the first $s(\rho_1)$ and the last $f(\rho_1)$ nodes, and ρ_2 connects C_2 to a terminal node and has no cyclic nodes with the exception of the first node $s(\rho_2)$. In our notation, $\gamma_0 = w(\rho_0)$, $\alpha_1 = w(C_1, f(\rho_0))$, $\alpha_2 = w(C_2, f(\rho_1))$, and, for $i = 1, 2$, $\gamma_i = w(\pi_i)w(\rho_i)$, where π_i is the directed path from $f(\rho_{i-1})$ to $s(\rho_i)$ taken along the cycle C_i.

We now describe a polynomial-time algorithm Φ_0 which, for a given simple cycle C_1 checks if there exists a path ρ_0 which connects q_0 to C_1, has no cyclic nodes with the exception of the last node $f(\rho_0)$, and for which $w(\rho_0)$ is not a left extension of a power of $w(C_1, f(\rho_0))$. To this end, for an arbitrary node c of C_1, we verify if there exists a directed path ρ which connects q_0 to c, has no cyclic nodes with the exception of c, and for which $w(\rho)$ is not a left extension of a power of $w(C_1, c)$. If q_0 is a cyclic node, then there is no such a path. Let q_0 be an acyclic node, and n be the number of acyclic nodes in G. It is clear that the length of ρ cannot be greater than n. For $j = 1, \ldots, n$, we can construct in a polynomial time a graph H^j satisfying the following conditions:

- All nodes of H^j are divided into $j+1$ sets $H_0^j, \ldots, H_j^j$ such that H_i^j, $1 \le i \le j-1$, contains all acyclic nodes v of G for which there is a directed path of the length i from q_0 to v without cyclic nodes, and there is a directed path of the length $j-i$ from v to c with only one cyclic node c;
- The set H_0^j contains only node q_0, and the set H_j^j contains only node c;
- For $i = 0, \ldots, j-1$, all edges which start in nodes belonging to H_i^j finish in nodes belonging to H_{i+1}^j;

- For $i = 0, \ldots, j-1$, any $u \in H_i^j$ and any $v \in H_{i+1}^j$, there is an edge from v to u labeled with a letter a if and only if in the graph G there is an edge from v to u labeled with the letter a.

One can show that the set of words corresponding to paths in H^j from q_0 to c is equal to the set of words corresponding to paths in G of the length j from q_0 to c.

For $j = 1, \ldots, n$, we construct in polynomial time a word $\beta^j = \beta_1^j \ldots \beta_j^j$ of the length j such that the word $\beta^j w(C_1, c)$ is a periodic word with period $|w(C_1, c)|$. Using Lemma 1 one can show that the considered path ρ does not exist if and only if, for $j = 1, \ldots, n$ and $i = 1, \ldots, j$, all edges between H_{i-1}^j and H_i^j are labeled with the letter β_i^j.

It is clear that there exists a polynomial-time algorithm Φ_1 which verifies if there is a directed path ρ_1 which connects C_1 to C_2 and has no cyclic nodes with the exception of the first and the last ones.

According to Observation 2, γ_2 is not a right extension of a power of α_2 if and only if the length of the path ρ_2 is at least 1. It means that we should verify if there exists a path which connects C_2 to an acyclic terminal node. It is clear that there exists a polynomial-time algorithm Φ_2 which solves this problem.

Using algorithms Φ_0, Φ_1 and Φ_2, we can check in polynomial time if there exists a metapath with exactly two cycles satisfying the conditions of Proposition 3. If such a metapath exists, then, by Proposition 3, $L(\mathcal{A})$ contains an infinite antichain.

If a metapath with exactly two cycles satisfying the conditions of Proposition 3 does not exist, we need to check if G contains a metapath τ with exactly one cycle satisfying the conditions of Proposition 2: $L(\tau) = \{\gamma_0 \alpha_1^{j_1} \gamma_1 : j_1 = 0, 1, \ldots\}$ and γ_0 is not a left extension of a power of α_1, and γ_1 is not a right extension of a power of α_1. We can do it in polynomial time using algorithms Φ_0 and Φ_2. If such a metapath exists, then, by Proposition 2, $L(\mathcal{A})$ contains an infinite antichain. Otherwise, $L(\mathcal{A})$ does not contain an infinite antichain. $\square$

Remark: in the proof of Theorem 2 we knowingly avoid estimating exact time complexity of the proposed solution, as this question is irrelevant for the purpose of this paper.

5 An alternative approach

In the attempt to extend the solution for languages to other combinatorial structures (permutations, graphs, etc.), in this section we propose an alternative approach to the same problem, which is not based on the notion of an automaton. We discuss possible ways to apply this approach to graphs and permutations in the last section.

The alternative approach is based on the notion of a *periodic* infinite antichain, which can be defined as follows.

Definition 1. An infinite antichain of words is periodic of period p if it has the form $\{\beta\alpha^k\gamma \mid k \in \mathbb{N}\}$, where $|\alpha| = p$ and neither β is a left extension of a power of α nor γ is a right extension of a power of α.

Notice that Lemma 3 verifies that the set $\{\beta\alpha^k\gamma \mid k \in \mathbb{N}\}$ is an antichain indeed. The notion of periodic infinite antichains and the results of Section 4 allow us to derive the following criterion of well-quasi-orderability of factorial languages defined by finitely many forbidden factors.

Theorem 3. *Let $D = \{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ be a finite set of pairwise incomparable words and $X = Free(\alpha_1, \ldots, \alpha_k)$ be the factorial language with the antidictionary D. Then X is well-quasi-ordered by the factor containment relation if and only if it contains no periodic infinite antichains of period at most $|\alpha_1| + |\alpha_2| + \ldots + |\alpha_k| + 1$.*

Proof. If X is well-quasi-ordered, then it certainly does not contain any periodic infinite antichains.

Conversely, suppose X is not well-quasi-ordered. From the construction of the automaton for factorial languages given in [6] we know that the the number of nodes in the automaton is precisely the number of different prefixes of the forbidden words. Hence the size of the automaton is at most $t = |\alpha_1| + |\alpha_2| + \ldots + |\alpha_k| + 1$.

Now, if the automaton contains two cycles C_1 and C_2 intersecting at some vertex v, Proposition 1 shows that the automaton contains an infinite antichain $\{\beta^{|C_1|}\alpha^i\beta \mid i \in \mathbb{N}\}$, where $\alpha = w(C_1, v)$ and $\beta = w(C_2, v)$. Hence X contains a periodic infinite antichain of period $|\alpha| = |C_1| \leq t$.

Suppose now the automaton contains a metapath with at least three cycles. Then by Proposition 4 the set $\{\gamma_0\alpha_1^{|\alpha_2|}\gamma_1\alpha_2^i\gamma_2 \mid i \in \mathbb{N}\}$ is a periodic infinite antichain of period $|\alpha_2| \leq t$.

Consider now the case when the automaton contains neither two intersecting cycles nor a metapath with at least three cycles. Then, as X is not well-quasi-ordered, from Propositions 2 and 3 we conclude that X contains one of the following antichains $\{\gamma_0\alpha_1^i\gamma_1 \mid i \in \mathbb{N}\}$ or $\{\gamma_0\alpha_1^i\gamma_1\gamma_2 \mid i \in \mathbb{N}\}$ or $\{\gamma_0\alpha_1^{|\alpha_2|}\gamma_1\alpha_2^i\gamma_2 \mid i \in \mathbb{N}\}$. These are again periodic infinite antichains of period either $|\alpha_1| \leq t$ or $|\alpha_2| \leq t$. Hence we conclude that if X is not well-quasi-ordered, then it contains a periodic antichain of period at most t. This completes the proof. □

The importance of this result is due to the fact that it suggests possible ways to approach the question of well-quasi-orderability for other combinatorial structures, such as graphs or permutations. We discuss this idea in Section 7.

Theorem 3 also raises an interesting question of determining the minimum value of the period in periodic infinite antichains that have to be broken to ensure well-quasi-orderability. According to the theorem, this value is at most $|\alpha_1| + |\alpha_2| + \ldots + |\alpha_k| + 1$. We believe that this bound can be substantially improved and discuss this question in the next section for factorial languages with binary alphabet.

6 Deciding WQO for factorial languages with binary alphabet

Throughout this section we deal with words in the binary alphabet $A = \{0, 1\}$. If a word β is not a left extension of a power of a word α, then we say that β is a *minimal* word with this property if any proper suffix of β is a left extension of a power of α. Similarly, we define the notion of a minimal word which is not a right extension of a power α. It is not difficult to see that if β is a minimal not left/right extension of a power of α, then $|\beta| \leq |\alpha|$.

The notions of minimal not left/right extensions allow us to define the notion of a minimal periodic infinite antichain as follows.

Definition 2. A periodic infinite antichain $\{\beta\alpha^k\gamma \mid k \in \mathbb{N}\}$ is minimal if β is a minimal not left extension of a power of α and γ is a minimal not right extension of a power of α.

To illustrate this notion, we list all minimal periodic infinite antichains of period at most 3 in Tables 1 and 2.

	Period 1		Period 2
(1.1)	0 1111 . . . 1111 0	(2.1)	00 1010 . . . 0101 00
(1.2)	1 0000 . . . 0000 1	(2.2)	11 0101 . . . 0101 00
		(2.3)	00 1010 . . . 1010 11
		(2.4)	11 0101 . . . 1010 11

Table 1: Minimal periodic infinite antichains of period 1 and 2

(3.1)	111 011011 . . . 110110 111	(3.10)	000 100100 . . . 001001 000
(3.2)	010 110110 . . . 110110 111	(3.11)	101 001001 . . . 001001 000
(3.3)	00 110110 . . . 110110 111	(3.12)	11 001001 . . . 001001 000
(3.4)	111 011011 . . . 011011 010	(3.13)	000 100100 . . . 100100 101
(3.5)	010 110110 . . . 011011 010	(3.14)	101 001001 . . . 100100 101
(3.6)	00 110110 . . . 011011 010	(3.15)	11 001001 . . . 100100 101
(3.7)	111 011011 . . . 011011 00	(3.16)	000 100100 . . . 100100 11
(3.8)	010 110110 . . . 011011 00	(3.17)	101 001001 . . . 100100 11
(3.9)	00 110110 . . . 011011 00	(3.18)	11 001001 . . . 100100 11

Table 2: Minimal periodic infinite antichains of period 3

We will say that a forbidden word α destroys an infinite antichain if α is a factor of all members of the antichain, except possibly finitely many of them. Theorem 3 tells us that in order to ensure well-quasi-orderability of a factorial language $Free(\alpha_1, \ldots, \alpha_k)$ it suffices to destroy periodic infinite antichains of period at most $|\alpha_1| + |\alpha_2| + \ldots + |\alpha_k| + 1$. We believe that in case of binary words we can do much better and conjecture the following.

Conjecture 1. Let Z be a finite set of words. Then the language $Free(Z)$ is well-quasi-ordered by the factor containment relation if and only if it does not contain minimal periodic infinite antichains of period at most $|Z|$.

Below we verify this conjecture for factorial languages with at most 3 forbidden words.

Lemma 7. *Let α be a binary word. Then $Free(\alpha)$ is well-quasi-ordered by the factor containment relation if and only if it does not contain minimal periodic infinite antichains of period 1.*

Proof. If the set $X = Free(\alpha)$ contains a periodic infinite antichain of period 1, then it is not well-quasi-ordered. So assume X does not contain minimal periodic antichains of period 1. As X does not contain the antichain (1.1), α must belong to the set $E_{1.1} = \{01^k, 1^k, 1^k0 \mid k \in \mathbb{N}\}$. Similarly, as X does not contain (1.2), α must belong to the set $E_{1.2} = \{10^k, 0^k, 0^k1 \mid k \in \mathbb{N}\}$. Therefore $\alpha \in E_{1.1} \bigcap E_{1.2} = \{0, 1, 01, 10\}$.

Clearly, the sets $Free(1) = \{0^k \mid k \in \mathbb{N}\}$ and $Free(0) = \{1^k \mid k \in \mathbb{N}\}$ are well-quasi-ordered. The sets $Free(01) = \{1^k0^l \mid k, l \in \mathbb{N}\}$ and $Free(10) = \{0^k1^l \mid k, l \in \mathbb{N}\}$ are well-quasi-ordered by Lemma 5. $\square$

Lemma 8. *Let α_1, α_2 be binary words. Then $Free(\alpha_1, \alpha_2)$ is well-quasi-ordered by the factor containment relation if and only if it does not contain minimal periodic infinite antichains of period at most 2.*

Proof. If the set $X = Free(\alpha_1, \alpha_2)$ contains a periodic infinite antichain of period at most 2, then it is not well-quasi-ordered. So assume X does not contain periodic infinite antichains of period at most 2. As X does not contain the antichain (1.1), one of α_1 and α_2 must belong to the set $E_{1.1} = \{01^k, 1^k, 1^k0 \mid k \in \mathbb{N}\}$. Similarly, as X does not contain the antichain (1.2), one of α_1 and α_2 must belong to the set $E_{1.2} = \{10^k, 0^k, 0^k1 \mid k \in \mathbb{N}\}$. Now either one of them belongs to $E_{1.1} \bigcap E_{1.2} = \{0, 1, 01, 10\}$, or we may assume, without loss of generality, that $\alpha_1 \in E_{1.1} \backslash E_{1.2} = \{01^k, 1^k, 1^k0 \mid k \geq 2\}$ and $\alpha_2 \in E_{1.2} \backslash E_{1.1} = \{10^k, 0^k, 0^k1 \mid k \geq 2\}$. In the latter case, α_2 has two consecutive 0's, so it does not destroy the antichain (2.4). Therefore, α_1 has to destroy (2.4) and in particular α_1 cannot contain three consecutive 1's, giving that $\alpha_1 \in \{11, 110, 011\}$. Similarly, we conclude that $\alpha_2 \in \{00, 100, 001\}$. We now notice that $Free(011, 001)$ contains the antichain (2.2) and $Free(110, 100)$ contains the antichain (2.3). Hence $Free(\alpha_1, \alpha_2)$ does not contain a periodic infinite antichain of period at most 2 only if one of the forbidden words belongs to $\{0, 1, 01, 10\}$ or

$$(\alpha_1, \alpha_2) \in \{(11, 00), (11, 100), (11, 001), (110, 00), (110, 001), (011, 00), (011, 100)\}.$$

If one of the forbidden words belongs to $\{0, 1, 01, 10\}$, then X is well-quasi-ordered by Lemma 7. The sets $Free(110, 001) = \langle\{(10)^k1^l, (01)^k0^l \mid k, l \in \mathbb{N}\}\rangle$ and $Free(011, 100) = \langle\{1^k(10)^l, 0^k(10)^l \mid k, l \in \mathbb{N}\}\rangle$ are well-quasi-ordered by Lemma 6. The remaining 5 sets not containing periodic antichains of period 2 are subsets of either $Free(110, 001)$ or $Free(011, 100)$, and hence are well-quasi-ordered too. $\square$

Lemma 9. *Let $\alpha_1, \alpha_2, \alpha_3$ be binary words. Then $Free(\alpha_1, \alpha_2, \alpha_3)$ is well-quasi-ordered by the factor containment relation if and only if it does not contain minimal periodic infinite antichains of period at most 3.*

Proof. If the set $X = Free(\alpha_1, \alpha_2, \alpha_3)$ contains a periodic infinite antichain of period at most 3, then it is not well-quasi-ordered. So assume X does not contain periodic infinite antichains of period at most 3. As X does not contain the antichain (1.1), one of the three forbidden words must belong to the set $E_{1.1} = \{01^k, 1^k, 1^k0 \mid k \in \mathbb{N}\}$. Similarly, as X does not contain the antichain (1.2), one of the three forbidden words must belong to the set $E_{1.2} = \{10^k, 0^k, 0^k1 \mid k \in \mathbb{N}\}$. So either one of the words belong to $E_{1.1} \bigcap E_{1.2} = \{0, 1, 01, 10\}$, or we may assume, without loss of generality, that $\alpha_1 \in E_{1.1} \backslash E_{1.2} = \{01^n, 1^n, 1^n0 \mid n \geq 2\}$ and $\alpha_2 \in E_{1.2} \backslash E_{1.1} = \{10^m, 0^m, 0^m1 \mid m \geq 2\}$, and hence the cases to analyze can be split into 9 groups:

(A)	$(\alpha_1, \alpha_2) \in \{(01^n, 0^m) \mid n, m \geq 2\}$	(E)	$(\alpha_1, \alpha_2) \in \{(1^n, 10^m) \mid n, m \geq 2\}$
(B)	$(\alpha_1, \alpha_2) \in \{(01^n, 10^m) \mid n, m \geq 2\}$	(F)	$(\alpha_1, \alpha_2) \in \{(1^n, 0^m1) \mid n, m \geq 2\}$
(C)	$(\alpha_1, \alpha_2) \in \{(01^n, 0^m1) \mid n, m \geq 2\}$	(G)	$(\alpha_1, \alpha_2) \in \{(1^n0, 0^m) \mid n, m \geq 2\}$
(D)	$(\alpha_1, \alpha_2) \in \{(1^n, 0^m) \mid n, m \geq 2\}$	(H)	$(\alpha_1, \alpha_2) \in \{(1^n0, 10^m) \mid n, m \geq 2\}$
		(I)	$(\alpha_1, \alpha_2) \in \{(1^n0, 0^m1) \mid n, m \geq 2\}$

Notice that the words in (E) can be obtained from the words in (A) by complementation (swapping 0's with 1's), the words in (F) can be obtained from the words in (A) by complementation followed by reversion (changing $w = v_1v_2 \dots v_n$ to $\bar{w} = v_nv_{n-1} \dots v_1$), the words in (G) can be obtained from the words in (A) by reversion, the words in (H) can be obtained from the words in (C) by complementation and the words in (I) can be obtained from the words in (B) by reversion. Therefore, we can restrict our attention to cases (A), (B), (C), (D).

Let $n = 2$ and $m = 2$. Then the words in (A), (B), (D) destroy all antichains. In case (C), the antichain (2.2) is not destroyed and hence α_3 has to destroy (2.2), i.e. α_3 must belong to $\langle\{11(01)^p, (01)^p00 \mid p \in \mathbb{N}\}\rangle$. So we obtain two triples $(011, 100, 11(01)^p)$ and $(011, 100, (01)^p00)$ and they destroy all periodic infinite antichains of period at most 3.

Let $n \geq 3$, $m \geq 3$, then $Free(\alpha_1, \alpha_2)$ contains antichains (2.1), (2.4), (3.9), (3.18). As α_3 destroys (2.1) and (2.4), α_3 must have zeros and ones alternating. Now, as α_3 destroys (3.9), the number of 0's in α_3 is at most 1 and to destroy (3.18), the number of 1's in α_3 must be at most one. We conclude that $\alpha_3 \in \{0, 1, 01, 10\}$, in which case X is well-quasi-ordered by Lemma 7.

For $n = 2$, $m \geq 3$ or $m = 2$, $n \geq 3$, we consider cases (A), (B), (C), (D) separately.

(A) If $n = 2, m \geq 3$, then $Free(011, 0^m)$ contains the antichains (2.1), (2.2), (3.13)-(3.15). So, to destroy (2.1) and (2.2), α_3 must belong to $\langle\{(01)^p00 \mid p \in \mathbb{N}\}\rangle$. But only $\alpha_3 = 0100$ and $\alpha_3 = 0101$ and their subwords destroy (3.13)-(3.15). Also notice that when $m > 3$, $Free(011, 0^m)$ contains (3.10)-(3.12), and only $\alpha_3 = 0100$ and its subwords destroy these antichains. If $n \geq 3$, $m = 2$, then $Free(01^n, 00)$ contains (2.4), (3.4), (3,5). So, to destroy (2.4), α_3 must belong to $\langle\{11(01)^p, (10)^p11 \mid p \in \mathbb{N}\}\rangle$. But only $\alpha_3 = 11010$ and $\alpha_3 = 1011$ and their subwords destroy (3.4) and (3.5). Also, when $n > 3$, $Free(01^n, 00)$ contains (3.1)-(3.3), and the maximal subword of 11010 destroying these antichains is 1101.

(B) If $n = 2$ and $m \geq 3$, then $Free(011, 10^m)$ contains (2.1), (2.2), (3.13)-(3.15). By case (A), to destroy these antichains, $\alpha_3 = 0100$ or $\alpha_3 = 0101$ or a subword of these. Also notice that $\alpha_3 = 0101$ can be only considered for $m = 3$, as $Free(011, 10000, 0101)$ contains (3.10)-(3.12). The triples for $n \geq 3$ and $m = 2$ follow from the previous paragraph by complementation.

(C) If $n = 2$ and $m \geq 3$, then $Free(011, 0^m1)$ contains (2.1), (2.2) and (3.12). Therefore, to destroy (2.1) and (2.2), α_3 must belong to $\langle\{(01)^p00 \mid p \in \mathbb{N}\}\rangle$ and only $\alpha_3 = 0100$ or its subwords destroy (3.12). The triples for $n \geq 3$ and $m = 2$ follow from the previous paragraph by reversion and complementation.

(D) If $n = 2$ and $m \geq 3$, then $Free(11, 0^m)$ contains the antichains (2.1) and (3.14). Therefore, to destroy (2.1), α_3 must belong to $\langle\{00(10)^k, (01)^k00 \mid k \in \mathbb{N}\}\rangle$. From these, only $00101, 10100$ and their subwords destroy (3.14). If $m > 3$, then $Free(11, 0^m)$ contains antichains (3.10)-(3.13). Therefore, to destroy (3.10), α_3 cannot contain an occurrence of 101, and α_3, given by the previous case, restricts to a subword of 0010 or 0100. The triples for $n \geq 3$, $m = 2$ follow from the previous paragraph by complementation.

Hence $Free(\alpha_1, \alpha_2, \alpha_3)$ does not contain a minimal periodic infinite antichain of period at most 3 if and only if

- either one of the forbidden words belongs to $\{0, 1, 01, 10\}$, in which case X is well-quasi-ordered by Lemma 7,
- or two of the forbidden words belong to

$$\{(11,00),(11,100),(11,001),(110,00),(110,001),(011,00),(011,100)\},$$

  in which case X is well-quasi-ordered by Lemma 8,
- or $(\alpha_1, \alpha_2, \alpha_3)$ is one of the triples (or complement or reversion of the triples) given in Table 3.

Case	$(\alpha_1, \alpha_2, \alpha_3)$	$Free(\alpha_1, \alpha_2, \alpha_3)$
(A)	$(011, 0^m, 0100)$	$\langle 1^k 0^r (10)^l \mid k, l, r \in \mathbb{N}, r < m\rangle$
(A)	$(011, 000, 0101)$	$\langle 1^k (001)^l, 1^k 0(100)^l \mid k, l \in \mathbb{N}\rangle$
(A)	$(01^n, 00, 1011)$	$\langle 01^r (01)^k, 1^k (01)^l \mid k, l, r \in \mathbb{N}, r < n\rangle$
(A)	$(0111, 00, 11010)$	$\langle 1^k (011)^l, (01)^k (011)^l \mid k, l \in \mathbb{N}\rangle$
(A)	$(01^n, 00, 1101)$	$\langle 1^k 0, (01)^k 1^r 0 \mid k, l, r \in \mathbb{N}, r < n\rangle$
(B)	$(011, 1000, 0101)$	$\langle 0^k (100)^l, 1^k 0(100)^l, 1^k (001)^l \mid k, l \in \mathbb{N}\rangle$
(B)	$(011, 10^m, 0100)$	$\langle 0^k (10)^l, 1^k 0^r (10)^l \mid \mid k, l, r \in \mathbb{N}, \quad r < m\rangle$
(C)	$(011, 001, 11(01)^p)$	$\langle 1^k (01)^r 0^l, (01)^k 0^l \mid k, l, r \in \mathbb{N}, r < p\rangle$
(C)	$(011, 001, (01)^p 00)$	$\langle 1^k (01)^r 0^l, 1^k (01)^l \mid k, l, r \in \mathbb{N}, r < p\rangle$
(C)	$(011, 0^m 1, 0100)$	$\langle 1^k 0^l, 1^k 0^r (10)^l \mid k, l, r \in \mathbb{N}, r < m\rangle$
(D)	$(11, 000, 00101)$	$\langle (10)^k (100)^l \mid k, l \in \mathbb{N}\rangle$
(D)	$(11, 000, 10100)$	$\langle (100)^k (10)^l \mid k, l \in \mathbb{N}\rangle$
(D)	$(11, 0^m, 0010)$	$\langle (01)^k 0^r 1 \mid k, l, r \in \mathbb{N}, r \leq m\rangle$
(D)	$(11, 0^m, 0100)$	$\langle 10^r (10)^k \mid k, l, r \in \mathbb{N}, r \leq m\rangle$

Table 3: Triples $(\alpha_1, \alpha_2, \alpha_3)$ destroying periodic infinite antichains of period at most 3

The first triple in the table gives the set $Free(011, 0^m, 0100)$, which is well-quasi-ordered because it consists of the union of m sets $\langle\{1^k 0^r (10)^l \mid k, l \in \mathbb{N}\}\rangle$, $r = 0, 1, \ldots, (m-1)$, each of which is well-quasi-ordered by Lemma 6. Similarly, all the sets in Table 3 can be easily seen as finite unions of sets which are well-quasi-ordered by Lemma 6. □

7 Beyond languages

The alternative solution to the problem of deciding well-quasi-orderability of factorial languages proposed in Section 5 suggests a possible way to approach the same problem for graphs and permutations. Similarly to languages, this approach is based on the notion of *periodic* infinite antichains and consists in checking the presence of antichains of only bounded periodicity. Below we outline this approach for the induced subgraph relation on graphs and briefly discuss it for the pattern containment relation on permutations.

To define the notion of a periodic infinite antichain for graphs, we use the notion of letter graphs introduced in [14] and slightly adapt it to our purposes. Originally, this notion was defined as follows.

Let A be a finite alphabet and $S \subseteq A^2$ a binary relation on A. With each word $\alpha = (\alpha_1\alpha_2\dots\alpha_n)$ over A we associate the graph G_α with vertices $\alpha_1, \alpha_2, \dots, \alpha_n$, in which two vertices α_i and α_j with $i < j$ are adjacent if and only if $(\alpha_i, \alpha_j) \in S$. If A consists of k letters, the graph G_α is called a k-letter graph. The importance of this notion for well-quasi-orderability is due to the fact that for each fixed k the set of all k-letter graphs is well-quasi-ordered by the induced subgraph relation [14].

We now modify the notion of letter graphs by distinguishing between consecutive and non-consecutive vertices of α. For nonconsecutive vertices α_i and α_j with $i < j$ the definition remains the same: α_i and α_j are adjacent if and only if $(\alpha_i, \alpha_j) \in S$. For consecutive vertices, we change the definition to the opposite: α_i and α_{i+1} are adjacent if and only if $(\alpha_i, \alpha_{i+1}) \notin S$. Let us denote the graph obtained in this way from the word α by G^*_α. For instance, if a is a letter of A and $(a, a) \notin S$, then the word $aaaaa$ defines a path on 5 vertices. With some restrictions, the induced subgraph relation on graphs defined in this way corresponds to the factor relation on words, i.e. G^*_α is an induced subgraph of G^*_β if and only if α is a factor of β.

The graph G^*_α constructed from a periodic word α will be called a periodic graph. The period of α will be called the period of G^*_α. In this way we define periodic graphs. To construct periodic antichains, we need to break the periodicity on both ends of the graphs (words). To this end, we simply color the first and the last vertices of the graph differently from the intermediate vertices (see Figure 2 for an illustration). If we now strengthen the induced subgraph relation by requiring that an embedding of one graph into another should respect the colors, then we convert the set of paths $P_3, P_4, P_5 \dots$ into an infinite periodic antichain of period 1.

Figure 2: A colored path on P_5

To justify this restriction to colored (also known as labelled) infinite antichain, let us observe that in [2] we conjecture that for hereditary classes of graphs defined by *finitely many* forbidden induced subgraphs, the notion of well-quasi-orderability by induced subgraphs coincides with the notion of well-quasi-orderability by labelled induced subgraphs and verify this conjecture for all known examples of well-quasi-ordered classes of graphs. Also, notice that according to the conjecture of Pouzet [15], in the case of labelled induced subgraphs one can be restricted to two different labels (colors).

For a finite collection C of graphs, let us denote by $t(C)$ the total number of vertices of graphs in C. Suggested by the result on languages, we conjecture the following.

Conjecture 2. There is a function $f : \mathbb{N} \to \mathbb{N}$ such that the class X of graphs defined by a finite collection C of forbidden induced subgraphs is well-quasi-ordered by the induced subgraph relation if and only if X contains no periodic infinite antichains of period at most $f(t(C))$.

To support the conjecture, let us notice that in the case of one forbidden induced subgraph G, the class $Free(G)$ is well-quasi-ordered if and only if it contains no colored periodic infinite antichains of period 1. Indeed, if G contains a cycle, then $Free(G)$ contains the antichain of 2-colored paths, and if G contains the complement of a cycle, then $Free(G)$ contains the antichain of complements of 2-colored paths. If G is free of cycles and their complements, then G is a path P_4 on 4 vertices (or its induced subgraph), in which case $Free(G)$ is well-quasi-ordered [7].

We also believe that a similar approach should also work for the pattern containment on permutations. Indeed, with each permutation $\pi : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ we can associate the permutation graph G_π on the vertex set $\{1, 2, \ldots, n\}$ in which two vertices i, j are adjacent if and only if $(i-j)(\pi(i)-\pi(j)) < 0$. Then the pattern containment on permutations corresponds to the induced subgraph relation on the permutation graphs.

References

[1] M. D. Atkinson, M. M. Murphy, and M. Ruškuc, Partially well-ordered closed sets of permutations, *Order*, 19 (2002) 101–113.

[2] A. Atminas, V.V. Lozin, Labelled induced subgraphs and well-quasi-ordering, *Order*, 32 (2015) 313–328.

[3] A. Atminas, V. Lozin, M. Moshkov, Deciding WQO for factorial languages, *Lecture Notes in Computer Science*, 7810 (2013) 68–79.

[4] R. Brignall, N. Ruškuc, and V. Vatter, Simple permutations: decidability and unavoidable substructures, *Theoretical Computer Science*, 391 (2008) 150–163.

[5] G.L. Cherlin and B.J. Latka, Minimal antichains in well-founded quasi-orders with an application to tournaments, *J. Combinatorial Theory B*, 80 (2000) 258–276.

[6] M. Crochemore, F. Mignosi, and A. Restivo, Automata and forbidden words, *Information Processing Letters*, 67 (1998) 111–117.

[7] P. Damaschke, Induced subgraphs and well-quasi-ordering, *J. Graph Theory*, 14 (1990) 427–435.

[8] A. de Luca and S. Varricchio, Well quasi-orders and regular languages, *Acta Informatica*, 31 (1994) 539–557.

[9] G. Ding, Subgraphs and well-quasi-ordering, *J. Graph Theory*, 16 (1992) 489–502.

[10] A. Finkel and Ph. Schnoebelen, Well-structured transition systems everywhere!, *Theoretical Computer Science*, 256 (2001) 63–92.

[11] N. Hine and J. Oxley, When excluding one matroid prevents infinite antichains, *Advances in Applied Mathematics*, 45 (2010) 74–76.

[12] N. Korpelainen and V.V. Lozin, Two forbidden induced subgraphs and well-quasi-ordering, *Discrete Mathematics*, 311 (2011) 1813–1822.

[13] J.B. Kruskal, The theory of well-quasi-ordering: a frequently discovered concept, *J. Combinatorial Theory A*, 13 (1972) 297–305.

[14] M. Petkovšek, Letter graphs and well-quasi-order by induced subgraphs, *Discrete Mathematics*, 244 (2002) 375–388.

[15] M. Pouzet, Un bel ordre d'abritement et ses rapports avec les bornes d'une multirelation, *C. R. Acad. Sci. Paris Sér. A-B*, 274 (1972) A1677–A1680.

[16] N. Robertson and P. Seymour, Graph Minors. XX. Wagner's conjecture, *J. Combinatorial Theory B*, 92 (2004) 325–357.

[17] D.A. Spielman and M. Bóna, An infinite antichain of permutations, *The Electronic J. Combinatorics*, 7 (2000) #N2.