

Institut für Informatik und Praktische Mathematik
der
Christian-Albrechts-Universität zu Kiel
D-24098 Kiel

Locally Threshold Testable Languages of Infinite Words

Thomas Wilke

Bericht Nr. 9203
Juli 1992

Locally Threshold Testable Languages of Infinite Words*

Thomas Wilke

Institut für Informatik und Praktische Mathematik

Christian-Albrechts-Universität zu Kiel

Hermann-Rodewald-Str. 3

D-W-2300 Kiel

Germany

E-mail: `tw@informatik.uni-kiel.dbp.de`

Abstract

Two versions of local threshold testability for languages of infinite words (ω -languages) are compared: It is proved that an ω -language is finitely locally threshold testable iff it is locally threshold testable and belongs to the Borel class $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$. As a consequence we obtain a result on the definability of infinite word structures in the signature of the successor function: It is decidable whether a given monadic second order formula has the same set of infinite word models as some first order formula. For biinfinite word models the corresponding problem was raised by JEAN ERIC PIN [LRS91, p. 172]. The major tool in the proofs is the analysis of De Bruijn graphs.

Introduction

The primary objective of this paper is to clarify the relationship between two versions of local threshold testability for languages of infinite words (ω -languages).

An ω -language L is called locally threshold testable if there are $l, m > 0$ such that the membership of a word α to L depends only on

- (a) the prefix of length $l - 1$ of α ,
- (b) the number of occurrences of factors of length l in α counted up to the threshold m (that means not distinguishing between numbers equal to or greater than m), and
- (c) the set of factors of length l occurring infinitely often.

*This paper was supported by ESPRIT Basic Research Action Working Group No. 3166 'Algebraic and Syntactic Methods in Computer Science' (ASMICS).

The language L is called finitely locally threshold testable if there are $l, m > 0$ such that (a) and (b) alone determine whether or not a given word α belongs to L . (It is also said that L is l -locally m -threshold or finitely l -locally m -threshold testable, respectively.)

Clearly every finitely locally threshold testable language is locally threshold testable, and every locally threshold testable language is regular (in the sense of recognizable by a Büchi automaton [Tho90, p. 136]).

The main result of this paper states that there is a certain topological condition characterizing the class of finitely locally threshold testable ω -languages within the larger class of locally threshold testable ω -languages. For that the set of all infinite words over a fixed alphabet is viewed as a topological space as usual (cf Section 1). Then it can be proved that an ω -language is finitely locally threshold testable iff it is locally threshold testable and belongs to the Borel class $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ that contains the sets which are at the same time countable unions of closed sets and countable intersections of open sets (Corollary 2(a)).

In view of the fact that both the class of regular ω -languages in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ and the class of locally threshold testable ω -languages are decidable [Lan69, Wil92], we obtain as a consequence: It is decidable whether a regular ω -language is finitely locally threshold testable (Corollary 2(b)).

Following J. RICHARD BÜCHI's tradition we use the last result (in formal language theory) to solve a problem in mathematical logic.

BÜCHI proved in [Büc62] that one obtains every regular ω -language over the alphabet $\{0, 1\}^k$ as the model set of a (closed) monadic second order formula (via an appropriate encoding), and vice versa. Moreover the transformations between Büchi automata and equivalent formulas were shown to be effective. WOLFGANG THOMAS proved in [Tho82] that the finitely locally threshold testable ω -languages correspond to the model sets of first order formulas.¹ Thus from the decidability of the property of a regular ω -language to be finitely locally threshold testable we obtain immediately that it is decidable whether for a given second order formula there is a first order formula with the same set of infinite word models (Corollary 5). – The corresponding result for finite word models was proved by DANIELE BEAUQUIER and JEAN ERIC PIN [BP91].

In the first section we fix the notation, give the basic definitions and state the main result. In the second section we discuss the aforementioned consequences in detail. The rest of the paper is dedicated to the proofs.

1 Notation and Main Result

Let A be an *alphabet*, i.e. a finite set. A *finite word* over A is a sequence $n \rightarrow A$ where $n = \{0, \dots, n-1\}$ is an initial segment of the set ω of natural numbers. An *infinite word* (ω -word) over A is a sequence $\omega \rightarrow A$. The domain of a word x is also called its length and is denoted by $|x|$.

We write A^* for the set of all finite words over A , use A^+ for the set of all finite and non-empty words over A , and A^ω is the notation for the set of all ω -words over A . The

¹Actually the paper gives the result for the case where the formulas are interpreted over finite word structures; but this result extends in an obvious way to structures over the set of natural numbers.

set of all words over A of length l is denoted by A^l . Similarly we write $A^{\geq l}$ for the set of all words of length greater than or equal to l .

Let x be an arbitrary word over A . For $i, j \in \omega \cup \{\omega\}$ the subsequence of x consisting of the items $x(k)$ with index $k \in \{l : i \leq l < j\}$ is denoted by $x(i, j)$; it is called a *factor* of x . The set of all *factors* of a fixed length $l \in \omega$ is defined by $\text{fact}_l(x) = \{u \in A^l \mid (\exists i) x(i, i+l) = u\}$. The set of all factors of length l occurring infinitely often in an infinite word α is defined accordingly: $\text{inf}_l(\alpha) = \{u \in A^l \mid (\forall i)(\exists j)(i < j \wedge \alpha(j, j+l) = u)\}$.

For $l \in \omega$ the l -*prefix* of a word x is defined by $\text{pref}_l(x) = x(0, l)$. The l -*suffix* of a finite word u of length greater than or equal to l is given by $u(|u| - l, |u|)$. If the length of u is less than l the word itself is its l -suffix. If α is an infinite word, every factor of the form $\alpha(i, \omega)$ is called a *suffix* of α .

Let \mathbb{N}_∞ denote the set \mathbb{N} of natural numbers augmented by ∞ . Throughout $\text{card}(X)$ denotes the power of the set X if it is finite, and ∞ otherwise.

Let x be an arbitrary word. The number of occurrences of a finite word u in x as a factor is defined by $\|x\|_u = \text{card}(\{i \mid x(i, i+l) = u\})$. The number of occurrences of a letter a in x is given by $|x|_a = \text{card}(\{i \mid a = x(i)\})$. For a set B of letters, we use $|x|_B$ as an abbreviation for $\sum_{a \in B} |x|_a$.

Given a natural number $m > 0$, we define the function $[\cdot]_m$ by

$$[\cdot]_m : \mathbb{N}_\infty \rightarrow m + 1$$

$$n \mapsto \begin{cases} n & \text{if } n < m, \\ m & \text{otherwise.} \end{cases}$$

With $m, l > 0$ and a word x we associate the function $\mu_x^{l,m}$ counting factors of length l up to the threshold m :

$$\mu_x^{l,m} : A^l \rightarrow m + 1$$

$$u \mapsto [\|x\|_u]_m.$$

Infinite words α and β are l -*locally m -threshold equivalent*, we write $\alpha \approx_\omega^{l,m} \beta$, if $\text{pref}_{l-1}(\alpha) = \text{pref}_{l-1}(\beta)$ and $\mu_\alpha^{l,m} = \mu_\beta^{l,m}$. They are *finitely l -locally m -threshold equivalent*, we write $\alpha \approx^{l,m} \beta$, if in addition $\text{inf}_l(\alpha) = \text{inf}_l(\beta)$.

An ω -language is called l -*locally m -threshold testable* if it is a union of $\approx_\omega^{l,m}$ -classes. It is *finitely l -locally m -threshold testable* if it is a union of $\approx^{l,m}$ -classes. It is called *[finitely] l -locally threshold testable* if it is [finitely] l -locally m -threshold testable for some m , and it is called *[finitely] locally threshold testable* if it is [finitely] l -locally threshold testable for some l .

The corresponding classes of regular ω -languages are denoted by $\mathcal{L}tt_{l,m}$, $\mathcal{L}tt_l$, $\mathcal{L}tt$ and *fin*- $\mathcal{L}tt_{l,m}$, *fin*- $\mathcal{L}tt_l$, *fin*- $\mathcal{L}tt$.

Concerning topologically defined ω -languages we start at the lowest level of the Borel hierarchy. An ω -language is called *open* if it has the form UA^ω for a subset U of A^* . It is a countable intersection of open sets if it has the form $\lim(U)$ for some subset U of A^* [Lan69, p. 378, Lemma 2.2], where $\alpha \in \lim(U)$ if for all i there is a $j > i$ such that $\alpha(0, j) \in U$. The corresponding set of languages is denoted by \mathcal{G}_δ . An ω -language is a countable union of closed sets if it is a complement of a language in \mathcal{G}_δ . The corresponding set of languages is denoted by \mathcal{F}_σ . (For details see [Tho90, pp. 152–156].)

In these terms we can state our main result.

Theorem 1 *We have*

$$(1) \quad \mathcal{L}tt_{l,m} \cap \mathcal{F}_\sigma \cap \mathcal{G}_\delta \subseteq \text{fin-}\mathcal{L}tt_{l,c}, \quad \text{for } l, m > 0 \text{ and } c = 2m \cdot \text{card}(A)^{2l}.$$

(That means, every l -locally m -threshold testable ω -language which belongs to $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ is finitely l -locally c -threshold testable.)

The proof occupies Sections 4 to 9. A sketch of it is given in Section 3.

How can the theorem be put in simple words? – The property of a locally threshold testable language to be in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ restricts the possible combinations of sets of factors occurring infinitely often in the words of the language. This restriction is even as strong that already the information about the number of occurrences of the factors counted up to a high finite threshold determines which factors occur infinitely often and which don't.

2 Consequences

2.1 Formal Language Theory

Corollary 2 (a) *We have*

$$\begin{aligned} \mathcal{L}tt \cap \mathcal{F}_\sigma \cap \mathcal{G}_\delta &= \text{fin-}\mathcal{L}tt, \\ \mathcal{L}tt_l \cap \mathcal{F}_\sigma \cap \mathcal{G}_\delta &= \text{fin-}\mathcal{L}tt_l, \quad \text{for } l > 0. \end{aligned}$$

(b) *It is decidable whether a regular ω -language is finitely locally threshold testable.*

Proof. (a) Both equations follow from Theorem 1 together with the inclusion $\text{fin-}\mathcal{L}tt \subseteq \mathcal{F}_\sigma \cap \mathcal{G}_\delta$ proved in [Wil92, Lemma 7.5, p. 43].

(b) This is a consequence of (a) and the fact that the properties of a regular ω -language to belong to $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ and to be locally threshold testable are decidable [Lan69, p. 279, Thm. 4.3],[Wil92, Corollary 7.8, p. 44].□

2.2 Mathematical Logic

An *infinite k -word structure* is a tuple $(\omega, \text{succ}, P_1, \dots, P_k)$ consisting of the set ω of natural numbers together with the successor function succ defined by $\text{succ}(n) = n + 1$ and k unary predicates P_1, \dots, P_k .

The set of closed monadic second order formulas using succ and P_1, \dots, P_k is denoted by S1S_k (second order with **1** successor over **k** predicates). Each interpretation $\mathbf{M} = (M_1, \dots, M_k)$ for the unary predicates of a formula belonging to S1S_k can be coded as an ω -word $w(\mathbf{M})$ over the alphabet 2^k in a natural way: the i -th component of the j -th letter of $w(\mathbf{M})$ is 1 iff $j \in M_i$. With $\phi \in \text{S1S}_k$ we associate the language $L(\phi)$ defined by $L(\phi) = \{\alpha \in 2^\omega \mid (\omega, \text{succ}, w(\mathbf{M})) \models \phi\}$.

As stated in the introduction the following is known by BÜCHI.

Theorem 3 [Büc62, pp. 5–7]

- (a) An ω -language $L \subseteq (2^k)^\omega$ is regular iff there is a formula $\phi \in \text{S1S}_k$ such that $L = L(\phi)$.
- (b) There is an algorithm computing for every k and $\phi \in \text{S1S}_k$ a BÜCHI automaton recognizing $L(\phi)$.

By F1S_k we denote the first order formulas of S1S_k . Their model sets were characterized by THOMAS.

Theorem 4 [Tho82, Thm. 4.2, p. 373] *An ω -language $L \subseteq (2^k)^\omega$ is finitely locally threshold testable iff there is a formula $\phi \in \text{F1S}_k$ such that $L = L(\phi)$.*

Theorem 3, Theorem 4 and Corollary 2 jointly yield the desired decidability result.

Corollary 5 *It is decidable whether a given formula $\phi \in \text{S1S}_k$ is equivalent to a formula $\psi \in \text{F1S}_k$ (in the sense that $L(\phi) = L(\psi)$).□*

3 Sketch of the Proof of Theorem 1

To verify the inclusion

$$(1) \quad \mathcal{L}tt_{l,m} \cap \mathcal{F}_\sigma \cap \mathcal{G}_\delta \subseteq \text{fin-}\mathcal{L}tt_{l,c}, \quad \text{for } l, m > 0 \text{ and } c = 2m \cdot \text{card}(A)^{2l},$$

of Theorem 1 we proceed in three major steps.

In the first step (Sections 4 and 5) we will investigate the properties of locally threshold testable languages which belong to $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$. We will join $\approx_\omega^{l,m}$ -classes into so-called l - m -blocks of largest possible size so that a given l -locally m -threshold testable language in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ is a union of these blocks. This reduces the original problem to the proof of the claim that every l - m -block is finitely l -locally c -threshold testable.

In the second step (Section 6) we will investigate a non-trivial example presented by the use of De Bruijn graphs. The l - m -blocks of this example will turn out to be not finitely l -locally m -threshold testable; this shows why we have to increase the parameter for the threshold in inequality (1). On the other hand we will find that the considered l - m -blocks are finitely l -locally testable though.

In the third and final step (Sections 7, 8 and 9) we will complete the proof of Theorem 1 by showing the claim which was set up at the end of the first step. We will essentially use what we will have learned from the example.

4 The \mathcal{G}_δ -Lemma

Lemma 6 *Let $L \in \mathcal{L}tt_{l,m} \cap \mathcal{G}_\delta$, $\alpha \in L$ and $\beta \approx_\omega^{l,m} \alpha$. If $\text{inf}_l(\alpha) \subseteq \text{inf}_l(\beta)$, then $\beta \in L$.*

Proof. Assume that $L = \lim(U)$. We are going to construct a word $\gamma \in \lim(U)$ such that $\gamma \approx_\omega^{l,m} \beta$, which is enough.

First of all notice that for every element u of $\text{inf}_l(\beta)$ there is a word $\gamma_u = uw_u\gamma_u'$ such that $\text{inf}_l(\gamma_u) = \text{inf}_l(\alpha)$ and $\text{fact}_l(\gamma_u) = \text{inf}_l(\beta) = \text{fact}_l(w_u)$.

Take a finite prefix u_0 of β (or α) such that $\text{pref}_{l-1}(u_0) = \text{pref}_{l-1}(\beta)$, $\mu_{u_0}^{l,m} = \mu_{\beta}^{l,m}$ and $\text{suff}_l(u_0) \in \text{inf}_l(\beta)$. Let $v_0 = \text{suff}_l(u_0)$. Since $u_0 w_{v_0} \gamma_{v_0} \in L$ by construction, there is a word u_1 belonging to U which is a prefix of $u_0 w_{v_0} \gamma_{v_0}$ such that $u_0 w_{v_0}$ in turn is a prefix of u_1 . Let $v_1 = \text{suff}_l(u_1)$. Then there is a word $u_2 \in U$ such that $u_1 < u_1 w_{v_1} < u_2$ by the same arguments as before. An iteration of this construction yields an increasing sequence

$$u_0 < u_0 w_{v_0} < u_1 < u_1 w_{v_1} < \dots \quad \text{with } \mu_{u_i}^{l,m} = \mu_{\beta}^{l,m} \text{ and } u_i \in U \text{ for } i \in \omega.$$

Since $\text{fact}_l(w_{v_i}) = \text{inf}_l(\beta)$ for $i \in \omega$ by construction, the limit γ of the sequence is the desired word belonging to $\lim(U)$ such that $\beta \approx_{\omega}^{l,m} \gamma$. \square

Corollary 7 *Let $L \in \mathcal{L}tt_{l,m} \cap \mathcal{F}_{\sigma}$, $\alpha \in L$ and $\beta \approx^{l,m} \alpha$. If $\text{inf}_l(\alpha) \supseteq \text{inf}_l(\beta)$, then $\beta \in L$.*

Proof. Assume that β does not belong to L . Then Lemma 6 applied to the complement $A^{\omega} \setminus L$ of L implies $\alpha \in A^{\omega} \setminus L$, which is a contradiction. (Notice that $\mathcal{L}tt_{l,m}$ is closed under complementation.) \square

5 De Bruijn graphs

By a *graph* $G = (V, E, i, t)$ we mean a set V of *vertices* and a set E of *edges* together with functions $i: E \rightarrow V$ and $t: E \rightarrow V$ determining *initial* and *terminal vertex*, respectively, of every edge.

The De Bruijn graph G_l of parameter $l > 0$ over the alphabet A is defined as follows (cf [Ber85, p. 237]).

- The set A^{l-1} of all words of length $l-1$ forms the set V of *vertices* of G_l .
- The set A^l of all words of length l forms the set E of *edges* of G_l .
- The *initial vertex* $i(e)$ of an edge e is its $l-1$ -prefix.
- The *terminal vertex* $t(e)$ of an edge e is its $l-1$ -suffix.

Every word x with length at least l gives us (exactly) one path through G_l :

$$\pi(x) = x(0, l), x(1, l+1), x(2, l+2), \dots,$$

and conversely every non-empty path determines a word with length at least l . We have the equation

$$|\pi(x)|_u = ||x||_u, \quad \text{for } x \in A^{\geq l}, u \in A^l,$$

where on the left hand side u is taken as a letter, i.e. as an edge, whereas on the right hand side u denotes a word.

Next we need the concepts of ‘induced subgraph’ and ‘connectedness’. As we are interested in the edge structure (rather than the vertex structure) of a given graph we define these as follows.

Let G be a graph as above. If $E' \subseteq E$ is a set of edges of G , the *subgraph* of G induced by E' is the quadruple $G[E'] = (V', E', i|_{E'}, t|_{E'})$ where a vertex v belongs to V' if there is an edge $e \in E'$ such that $i(e) = v$ or $t(e) = v$.

A path in G is a non-empty word $p: n \rightarrow E$ such that $t(e_j) = i(e_{j+1})$ for every $j < n$. If p is finite we say that the p connects e_0 with e_{n-1} . Edges e, e' are *connected* if there are paths p and p' connecting e with e' and e' with e , respectively. In this case we write $e \sim e'$. The *connectedness relation* \sim is transitive and symmetric but need not to be reflexive (in contrast to the usual definition). A set $E' \subseteq E$ of edges in G is said to be *connected* in G if $E' \times E'$ is a subset of \sim .

We observe that for any connected subset E' of E the set $\{e \in E \mid (\exists e' \in E') e \sim e'\}$ is the largest connected subset including E' . Its induced subgraph is called the *connected component* of E' . Obviously, edge and vertex set of distinct connected components are pairwise disjoint.

For the rest of the paper we restrict our considerations to De Bruijn graphs. We assume that the De Bruijn graph $G = G_l$ of parameter l over the given alphabet A is given as (V, E, i, t) . We are interested in the connected components C_1, \dots, C_n of the subgraph $G' = G[E']$ induced by a set of edges $E' \subseteq E$, which changes from case to case. We write E_i for the edge set of C_i and V_i for its vertex set. The union of all E_i is denoted by \tilde{E} , the set R is defined by $R = E' \setminus \tilde{E}$.

For a fixed E' we consider only words α such that $\text{inf}_l(\alpha) \subseteq E'$. With every such word we associate the connected component C_α of $\text{inf}_l(\alpha)$ in G' . We write E_α for the edge set and V_α for the vertex set of C_α .

Often the set E' will be given by a $\approx^{l,m}$ -class $K: E' = (\mu_\alpha^{l,m})^{-1}(m)$ for some $\alpha \in K$. This definition does not depend on α , so that we will simply write $E' = E_K$.

A first application of De Bruijn graphs is a reformulation of the \mathcal{G}_δ -lemma. Let K be a fixed $\approx^{l,m}$ -class and let $E' = E_K$ (see above).

Corollary 8 *If $L \in \mathcal{L}tt_{l,m} \cap \mathcal{F}_\sigma \cap \mathcal{G}_\delta$, $\alpha, \beta \in K$, and $E_\alpha = E_\beta$, then $\alpha \in L$ iff $\beta \in L$.*

Proof. By definition we have $E' = (\mu_\alpha^{l,m})^{-1}(m)$. Thus there is a word γ such that $\alpha \approx^{l,m} \beta \approx^{l,m} \gamma$ and $\text{inf}_l(\gamma) = E_\alpha$. Consequently $\alpha \in L$ iff $\gamma \in L$ by Lemma 6 together with Corollary 7, and $\beta \in L$ iff $\gamma \in L$ by the same arguments. \square

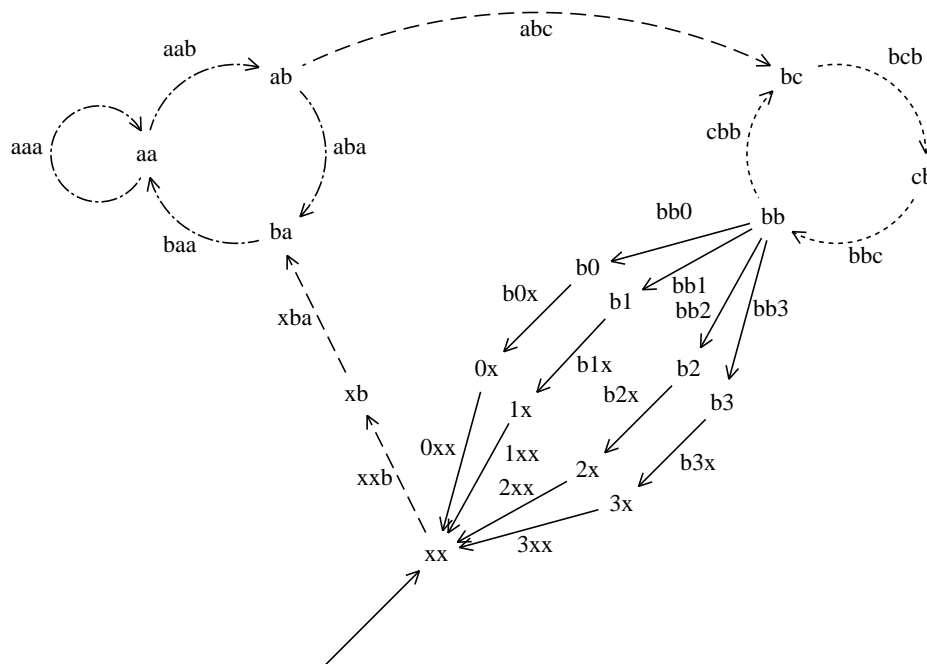
This result leads to the following definition: An *l - m -block* of a $\approx^{l,m}$ -class K is a maximal subset of K of words with the same associated connected component, i.e. it is a set of the form $\{\beta \in K: E_\alpha = E_\beta\}$ for some $\alpha \in K$. (Notice that the blocks of K partition K .)

In these terms every l -locally m -threshold testable language in $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ is a union of l - m -blocks. Thus for the desired proof of Theorem 1 it suffices to show the following claim.

Claim 9 *Every l - m -block is finitely l -locally c -threshold testable.*

6 Demonstration

We present the promised example. Consider Figure 1.

Figure 1: Graph H

It shows a graph H which is a subgraph of the DE BRUIJN graph of parameter 3 over the alphabet $A = \{a, b, c, x, 0, 1, 2, 3\}$. We define K to be the language which contains an ω -words $\alpha \in A^\omega$ iff

- $\pi(\alpha)$ is a path through H ,
- $\pi(\alpha)$ starts in xx ,
- $|\pi(\alpha)|_e = 1$ iff the edge e is drawn as a connected arrow,
- $|\pi(\alpha)|_e \geq 2$ iff the edge e is drawn as a dotted arrow.

Then K is finitely 3-locally 2-threshold testable since it is a $\approx^{l,m}$ -class.

Consider the connected components of $H' = H[E_K]$. There are exactly two of them: the first in the upper left corner of H , the second in the upper right corner. Thus we have two 3-2-blocks in K . The first, B_l , consists of the words in K which eventually end in the upper left corner of the graph, and the second, B_r , contains all words in K which eventually enter the upper right corner and stay there forever.

Our aim is to show that both blocks are finitely 3-locally threshold testable. Obviously they are not finitely 3-locally 2-threshold testable. Thus we must increase the threshold. But how far?

We want to distinguish between the cases that a word of K eventually stays in the left or right component of H' . Each path of a word of K passes exactly four times the

space between the upper right circle of \mathbb{H} and its origin xx , because every direct path from bb to xx must be used exactly once. Therefore, the edge abc is passed four or five times, and this depends just on which component the path finally stays in. This means: A word of K belongs to B_r iff its path has at least five occurrences of abc , otherwise it belongs to B_l . Thus B_l and B_r are finitely 3-locally 5-threshold testable. —

In the present example we saw that we can determine the block of a given word by threshold counting of edges coming in and going out of the connected components of the reduced graph G' . This illustrates a **general phenomenon** which we are going to explain in detail.

Let us fix an arbitrary $\approx^{l,m}$ -class K and let $E' = E_K$. Our aim is to determine the connected component C_α associated with an arbitrary word α in K , i.e. the l - m -block α belongs to.

Look at the path of α through G . Since $\text{inf}_l(\alpha) \subseteq C_\alpha$ there is either some position in the path where an edge occurs which goes into C_α and all edges after that position belong to C_α , or else all edges of the path belong to C_α right from the beginning.

In the latter case we have no edges coming in or going out of any component in G' , and the $l-1$ -prefix of α belongs to V_α . Thus we can determine C_α since the $l-1$ -prefix is the same for all words of a $\approx^{l,m}$ -class, in particular for all words of K .

In the former case we have this final incoming edge, and before that every edge coming into any component C_i matches an edge going out of it.

That means if we could count the number of occurrences of incoming and outgoing edges, we would be able to determine C_α , just by counting modulo 2 (Lemma 11): If all paths of words of K start outside any connected component, only for C_α there is an odd number of incoming and outgoing edges. If all paths start from the component C_i then $C_i = C_\alpha$ iff there is an even number of incoming and outgoing edges for C_i ; otherwise C_α is the component different from C_i for which we count an odd number of incoming and outgoing edges.

As a matter of fact we are able to precisely count incoming and outgoing edges, although we are only allowed to count up to a fixed finite threshold. In fact it can be proved that the number of incoming and outgoing edges is bounded for all paths of the words of an arbitrary class K (Corollary 13). If this number were not bounded, we could show that there is an incoming or outgoing edge which belongs to a connected component – a contradiction.

There is just one thing to mention before we go into details. Counting or pairing of incoming and outgoing edges is not that easy. It may happen that an edge is at the same time an incoming and outgoing edge for the same component or an incoming edge for one and an outgoing edge for another component. But this can also be fixed.

7 IN-/OUT-Counting

In this section E' is an arbitrary set of edges in the graph G_l .

The set of *incoming* edges (in G_l) of a connected component C_i of G' is defined by

$$\text{IN}_i = \{e \in E \setminus E_i : t(e) \in V_i\}.$$

The set of *outgoing* edges of C_i is defined analogously:

$$\text{OUT}_i = \{e \in E \setminus E_i : i(e) \in V_i\}.$$

The union of all sets IN_i is denoted by IN , the union of all sets OUT_i is denoted by OUT , and IO is defined as to be $\text{IN} \cup \text{OUT}$.

Remark 10 (a) The sets IN_i are pairwise disjoint.

(b) The sets OUT_i are pairwise disjoint.

(c) $\text{IO} \cap E' \subseteq R$.

In the remainder of this section we will analyze how incoming and outgoing edges occur in a path through G_l .

Lemma 11 *Let p be an arbitrary finite path through G_l starting with e and ending with f . Set $u = \text{pref}(e)$ and $v = \text{suff}(f)$. Denote by χ_i the characteristic function of V_i . Then*

$$|p|_{\text{IN}_i} + \chi_i(u) = |p|_{\text{OUT}_i} + \chi_i(v), \quad \text{for } i \in \{1, \dots, n\}.$$

Proof. The proof proceeds by induction on the length of p .

Induction base, $|p| = 1$. Assume p consists of the single edge g . We proceed by case distinction depending on the values $\chi_i(u)$ and $\chi_i(v)$.

Case $\chi_i(u) = \chi_i(v) = 0$. Then g is neither in OUT_i nor in IN_i . Thus the equation is true.

Case $\chi_i(u) = 1$ and $\chi_i(v) = 0$. Then g is in OUT_i but not in IN_i . Hence, the equation is true.

Case $\chi_i(u) = 0$ and $\chi_i(v) = 1$. Then g is in IN_i but not in OUT_i . Hence, the equation is true.

Case $\chi_i(u) = \chi_i(v) = 1$. Then neither $g \in \text{IN}_i$ nor $g \in \text{OUT}_i$, or $g \in \text{IN}_i \cap \text{OUT}_i$. Thus, the equation is also true.

Induction step, $|p'| = j + 1$. Assume the equation holds for p with length $\leq j$. Write p' as pg with $|p| = j$ and $g \in E$. Let u and v as in the formulation of the lemma, and let $u' = \text{pref}(g)$ and $v' = \text{suff}(g)$. Then by induction hypothesis

$$|p|_{\text{IN}_i} + \chi_i(u) = |p|_{\text{OUT}_i} + \chi_i(v) \quad \text{and} \quad |g|_{\text{IN}_i} + \chi_i(u') = |g|_{\text{OUT}_i} + \chi_i(v').$$

Adding up both equations yields

$$|p'|_{\text{IN}_i} + \chi_i(u) + \chi_i(u') = |p'|_{\text{OUT}_i} + \chi_i(v') + \chi_i(v).$$

Thus it remains to show $\chi_i(v) = \chi_i(u')$, but this is true, for $u' = v$ by definition. \square

8 α -Partitions

Throughout this section, $E' = E_K$ for some $\approx^{l,m}$ -class K . Our aim is to show that the value $|\pi(\alpha)|_{\text{IO}}$ is bounded for all $\alpha \in K$.

Lemma 12 *Let $\alpha \in K$.*

- (a) *If $p \in E'^*$ is a factor of $\pi(\alpha)$, then $|w|_R \leq \text{card}(A)^l$.*
- (b) $|\pi(\alpha)|_{A^l \setminus E'} \leq (m-1)\text{card}(A)^l$.

Proof. (a) Assume that the first inequality does not hold. Then there is an edge e of R which occurs (at least) twice in p . Thus e belongs to a component of G' ; hence $e \in \tilde{E}$ which contradicts $e \in R = E \setminus \tilde{E}$.

(b) Recall that every word of length l not belonging to E' occurs less than m times in α . \square

Corollary 13 *Let $\alpha \in K$. Then $|\pi(\alpha)|_{\text{IO}} < 2m \cdot \text{card}(A)^l$.*

Proof. Since $(\mu_\alpha^{l,m})^{-1}(m) = E'$, the path $\pi(\alpha)$ allows a decomposition as follows:

$$\pi(\alpha) = u_0 v_1 u_1 v_2 \dots v_r \pi',$$

where $u_i \in E'^+$, $v_i \in (A^l \setminus E')^+$ for $i > 0$, $u_0 \in E'^*$, $\pi' \in \tilde{E}^\omega$. We split $|\pi(\alpha)|_{\text{IO}}$ into $|u_0 u_1 \dots u_{r-1}|_{\text{IO}}$ and $|v_1 v_2 \dots v_r \pi'|_{\text{IO}}$. Since $\text{IO} \cap \tilde{E} = \emptyset$, the latter term is equal to $|v_1 v_2 \dots v_r|_{\text{IO}}$.

From Lemma 12(b) we obtain

$$(i) \quad r \leq (m-1)\text{card}(A)^l.$$

Then

$$\begin{aligned} |u_0 u_1 \dots u_{r-1}|_{\text{IO}} &\leq |u_0 u_1 \dots u_{r-1}|_R && \text{Remark 10} \\ &= \sum_{i=0}^{r-1} |u_i|_R \\ &\leq r \cdot \text{card}(A)^l && \text{Lemma 12(a)} \\ &\leq (m-1)\text{card}(A)^l \text{card}(A)^l && (i). \end{aligned}$$

On the other hand we have

$$\begin{aligned} |v_1 v_2 \dots v_r|_{\text{IO}} &\leq |v_1 v_2 \dots v_r|_{A^l \setminus E'} && v_i \in A^l \setminus E' \\ &\leq |\pi(\alpha)|_{A^l \setminus E'} \\ &\leq (m-1)\text{card}(A)^l. && \text{Lemma 12(a)} \end{aligned}$$

Adding up these inequalities yields

$$\begin{aligned} |\pi(\alpha)|_{\text{IO}} &\leq (m-1)\text{card}(A)^l \text{card}(A)^l + (m-1)\text{card}(A)^l \\ &< 2m \cdot \text{card}(A)^{2l}. \end{aligned}$$

\square

9 Proof of Claim 9 and Theorem 1

Remember that we reduced the proof of Theorem 1 to the proof of Claim 9 (Section 5): Every l - m -block is finitely l -locally c -threshold testable.

We will show that, for any $\approx^{l,c}$ -equivalent words α and β , if α belongs to the l - m -block B , then β belongs to B' .

Let α belong to the $\approx^{l,m}$ -class K . Consider $E' = E_K$. Since $\approx^{l,c} \subseteq \approx^{l,m}$, the word β belongs to an l - m -block B' in K . Then we are left with the proof of $B = B'$. This in turn is equivalent to $C_i = C_j$, when C_i is the component associated with α and C_j is the component associated with β .

Both paths $\pi(\alpha)$ and $\pi(\beta)$ allow decompositions $p\pi$ and $p'\pi'$ respectively such that p and p' are not empty and γ and γ' do not contain edges of IO. Furthermore we can assume that the last edge of p and every edge in π belong to E_i and, analogously, that the last edge of p' and every edge in π' belong to E_j .

Recall that c was defined just as to be the right hand side of the inequality in Corollary 13. Thus

$$(i) \quad |\alpha|_u = |\beta|_u, \quad \text{for } u \in \text{IO}.$$

Let u' denote the $l-1$ -prefix of the first edge of p' and v' the $l-1$ -suffix of the last edge of p' . We have $\chi_i(v) = 1$, thus Lemma 11 implies

$$(ii) \quad |p|_{\text{IN}_i} + \chi_i(u) = |p|_{\text{OUT}_i} + 1 \text{ and } |p'|_{\text{IN}_i} + \chi_i(u') = |p'|_{\text{OUT}_i} + \chi_i(v').$$

From (i) we may conclude

$$|p|_{\text{IN}_i} = |p'|_{\text{IN}_i} \text{ and } |p|_{\text{OUT}_i} = |p'|_{\text{OUT}_i}.$$

Thus (ii) yields $\chi_i(u) = 1$ and $\chi_i(u') = \chi_i(v')$. Since $u = \text{pref}_{l-1}(\alpha) = \text{pref}(\beta) = u'$ we finally obtain $\chi_i(v') = 1$, hence $i = j$, whence $B = B'$, as was to be shown.

This completes the proof of Claim 9 and Theorem 1.

Discussion

The fact that the class of finitely locally threshold testable ω -languages equals the intersection of the class of locally threshold testable languages with the Borel class $\mathcal{F}_\sigma \cap \mathcal{G}_\delta$ gives us a decision procedure for testing the property of being finitely locally threshold testable for a regular ω -language, which can also be used in the model theory of infinite word structures.

In the special case $m = 1$ it was shown [Wil92, Thm. 7.9, p. 45] that c can be chosen as 1. We leave open the question which is the optimal c in Theorem 1, and have as yet no results about the complexity of the involved decision procedures.

Acknowledgement

I am very grateful to MANFRED SCHIMMLER who provided me with a lot of helpful counter examples and to DOMINIQUE PERRIN who explained the notion of De Bruijn graphs ton of De Bruijn graphs to me.

References

- [Ber85] Claude Berge, *Graphs*, second revised ed., North-Holland Mathematical Library, vol. 6, North-Holland, Amsterdam, 1985, part 1.
- [BP91] Daniele Beauquier and Jean Eric Pin, *Languages and scanners*, Theoret. Comput. Sci. **84** (1991), no. 1, 3–21.
- [Büc62] J. Richard Büchi, *On a decision method in restricted second order arithmetic*, Logic, Methodology, and Philosophy of Science: Proc. of the 1960 International Congress (E. Nagel, P. Suppes, and A. Tarski, eds.), Stanford University Press, 1962, pp. 1–11.
- [Lan69] Laurence H. Landweber, *Decision problems for ω -automata*, Math. Systems Theory **3** (1969), 376–384.
- [LRS91] *Proc. workshop 'logics and recognizable sets'* (Wolfgang Thomas, ed.), Technical Report 9104, Inst. f. Informatik u. Prakt. Math., Univ. Kiel, Germany, 1991.
- [Tho82] Wolfgang Thomas, *Classifying regular events in symbolic logic*, J. Comput. System Sci. **25** (1982), 360–376.
- [Tho90] Wolfgang Thomas, *Automata on infinite objects*, Handbook of Theoretical Computer Science (Jan van Leeuwen, ed.), Elsevier Science Publishers B.V., 1990, pp. 134–191.
- [Wil92] Thomas Wilke, *An algebraic theory for regular languages of finite and infinite words*, Technical Report 9202, Inst. f. Inform. u. Prakt. Math., Univ. Kiel, Germany, 1992.