Well-quasi-orders and Regular ω-languages

Mizuhito Ogawa [a]

[a] *Japan Advanced Institute of Science and Technology*
1-1 Asahidai Tatsunokuchi Nomi Ishikawa 923-1292 Japan
mizuhito@jaist.ac.jp

Abstract

In "*On regularity of context-free languages, Theoretical Computer Science Vol.27, pp.311-332, 1983*", Ehrenfeucht et al. showed that a set L of finite words is regular if and only if L is $\leq$-closed under some monotone well-quasi-order (WQO) $\leq$ over finite words. We extend this result to regular ω-languages. That is,

(1) an ω-language L is regular if and only if L is $\preceq$-closed under a *periodic* extension $\preceq$ of some monotone WQO over finite words, and

(2) an ω-language L is regular if and only if L is $\preceq$-closed under a WQO $\preceq$ over ω-words that is a *continuous* extension of some monotone WQO over finite words.

Key words: ω-language, well-quasi-order, regularity.

1 Preliminaries

Throughout the paper, we will use A for a finite alphabet, A^* for a set of all (possibly empty) finite words on A, and A^ω for a set of all ω-words on A. A concatenation of two words u, v is denoted by u, v, an element-wise concatenation of two sets U, V of words by $U.V$, $\underbrace{V.V.\cdots.V}_{i}$ by V^i, and $V.V.V.\cdots$ by V^ω.

The length of a finite word u is denoted by $|u|$. As a convention, we will use ϵ for the empty word, $u, v, w, \cdots$ for finite words, $\alpha, \beta, \cdots$ for ω-words, $a_1, a_2, \cdots$ for elements in A, $i, j, k, l, \cdots$ for indices, and $U, V, \cdots$ (capital letters) for sets. We sometimes use $x, y, \cdots$ for elements of a set.

A regular ω-language is a set of ω-words that are accepted by a (nondeterministic) *Büchi* automaton $\mathcal{A} = \{Q, q_0, \Delta, F\}$, where Q is a finite set of states,

q_0 an initial state, $\Delta \subseteq Q \times A \times Q$ a transition relation, and F a set of final states. $\alpha = a_1 a_2 a_3 \cdots \in A^\omega$ is accepted by $\mathcal{A}$ if its corresponding run $q_0 \underset{a_1}{\rightarrow} q_1 \underset{a_2}{\rightarrow} q_2 \underset{a_3}{\rightarrow} \cdots$ runs through some state of F infinitely often. A set of ω-words accepted by $\mathcal{A}$ is denoted by $L(\mathcal{A})$. For states q, q' and $w \in A^*$, we write $q \underset{w}{\rightarrow} q'$ if there is a run of $\mathcal{A}$ on w, and we write $q \overset{F}{\underset{w}{\rightarrow}} q'$ if there is a run of $\mathcal{A}$ on w from q to q' such that the run runs through some state of F.

A congruence $\sim$ is an equivalent relation over A^* preserved by concatenations. A congruence $\sim$ is finite if there are only finitely many $\sim$-classes. Details are given elsewhere [3].

Definition 1.1 Let $L \subseteq A^\omega$ and let $\sim$ be a congruence over A^*. We say that $\sim$ saturates L if for each $\sim$-class U, V, $U.V^\omega \cap L \neq \phi$ implies $U.V^\omega \subseteq L$.

Lemma 1.2 For a *Büchi* automaton $\mathcal{A}$ and $u, v \in A^*$, we define $u \sim_{\mathcal{A}} v$ if $(q \underset{u}{\rightarrow} q' \Leftrightarrow q \underset{v}{\rightarrow} q') \ \wedge \ (q \overset{F}{\underset{u}{\rightarrow}} q' \Leftrightarrow q \overset{F}{\underset{v}{\rightarrow}} q')$ for each $q, q' \in Q$. Then $\sim_{\mathcal{A}}$ is a finite congruence that saturates $L(\mathcal{A})$.

Theorem 1.3 $L \subseteq A^\omega$ is regular if and only if some finite congruence saturates L.

Lemma 1.4 Let $\sim$ be a finite congruence over A^*.

(1) Let $\alpha = u_1 u_2 \cdots \in A^\omega$ and let $u(i,j) = u_i u_{i+1} \cdots u_{j-1}$ where $u_i \in A^*$. There exist a $\sim$-class V and $i_1 < i_2 < \cdots$ such that $u(i_j, i_k) \in V$ for each j, k with $j < k$.
(2) Let U, V be $\sim$-classes. There exist $\sim$-classes U', V' such that $U.V^\omega \subseteq U'.V'^\omega$, $U'.V' \subseteq U'$, and $V'.V' \subseteq V'$.

Proof

(1) Since $\sim$ has only finitely many $\sim$-classes, this is a direct consequence of (infinite) *Ramsey Theorem.*
(2) Note that for each $\sim$-class $U_1, \cdots, U_m, W$, $U_1. \cdots .U_n \cap W \neq \emptyset$ implies $U_1. \cdots .U_n \subseteq W$. Since $\sim$ has only finitely many $\sim$-classes, from (infinite) *Ramsey Theorem* there exist a $\sim$-class V' and $i_1 < i_2 < \cdots$ such that $V^{i_k - i_j} \subseteq V'$ for each j, k with $j < k$ and $V'.V' \subseteq V'$. Let U' be a $\sim$-class that includes $U.V^{i_1}$. Then $U.V^\omega \subseteq U'.V'^\omega$, $U'.V' \subseteq U'$, and $V'.V' \subseteq V'$. ■

We denote a quasi-order (QO, i.e., reflexive transitive binary relation) over a set S by $(S, \leq)$. If S is clear from the context, we simply denote by $\leq$. As a convention, a QO over finite words is denoted it by $\leq$, and a QO over ω-words is denoted by $\preceq$.

Definition 1.5 For a QO $(S, \leq)$ and $L \subseteq S$, L is $\leq$-closed if for each $x \in L$ $x \leq y$ implies $y \in L$.

Definition 1.6 A QO $(S, \leq)$ is a well-quasi-order (WQO) if for any infinite sequence $x_1, x_2, \cdots$ in S, there exist i, j such that $i < j$ and $x_i \leq x_j$.

A QO $(A^*, \leq)$ is monotone if $u \leq v$ implies $w_1 u w_2 \leq w_1 v w_2$ for each $u, v, w_1, w_2 \in A^*$.

2 First theorem

Definition 2.1 A QO $(A^\omega, \preceq)$ is a periodic extension of $(A^*, \leq)$ if the following conditions are satisfied:

- For each $u_i, v_i \in A^*$, $u_i \leq v_i$ for any i implies $u_1 u_2 u_3 \cdots \preceq v_1 v_2 v_3 \cdots$.
- For each $\alpha \in A^\omega$, there exist $u, v \in A^*$ such that $\alpha \preceq u.v^\omega$ and $\alpha \succeq u.v^\omega$.

Theorem 2.2 Let $L \subseteq A^\omega$. L is regular if and only if L is $\preceq$-closed under a periodic extension $(A^\omega, \preceq)$ of a monotone WQO $(A^*, \leq)$.

For instance, the embedding over ω-words is the periodic extension of the embedding over finite words. Note that a periodic extension of a monotone WQO over A^* is a WQO over A^ω. We will prove Theorem 2.2 below.

Lemma 2.3 Let $\sim$ be a finite congruence on A^* and let U, V be $\sim$-classes. For $u, v \in A^*$, if $uv^\omega \in U.V^\omega$, $U.V \subseteq U$, and $V.V \subseteq V$, there exist $w_1 \in U$ and $w_2 \in V$ such that $w_1 w_2^\omega = uv^\omega$.

Proof Let $uv^\omega = u'v'_1v'_2 \cdots$ satisfying $u' \in U$ and $v'_i \in V$, and let $w(i,j) = v'_i \cdots v'_{j-1}$ for $i < j$. Let $k_j \equiv |w(1,j)|\ (mod\ |v|)$. Then there exist k_{j_1} and k_{j_2} such that $k_{j_1} < k_{j_2}$ and $k_{j_1} \equiv k_{j_2}\ (mod\ |v|)$. Since there are infinitely many such pairs, we can assume that $|u| \leq |u'w(1, j_1 - 1)|$. Let $w_1 = u'.w(1, j_1 - 1)$ and $w_2 = w(j_1, j_2 - 1)$. Since $U.V \subseteq U$ and $V.V \subseteq V$, $w_1 \in U$, $w_2 \in V$ and $uv^\omega = w_1 w_2^\omega$. ■

Lemma 2.4 For a *Büchi* automaton $\mathcal{A}$ and $\alpha \in A^\omega$, let $[\![\alpha]\!] = \{U.V^\omega \mid \alpha \in U.V^\omega\}$ where U, V are $\sim_\mathcal{A}$-classes. We define $\alpha \preceq' \beta$ if $[\![\alpha]\!] \cap [\![\beta]\!] \neq \emptyset$. Then,

(1) $L(\mathcal{A})$ is $\preceq'$-closed.
(2) $u_i \sim_\mathcal{A} v_i$ for each i imply $u_1 u_2 \cdots \preceq' v_1 v_2 \cdots$.

Proof From Lemma 1.2, $\sim_\mathcal{A}$ saturates L and $U.V^\omega \subseteq L$ for each $U.V^\omega \in [\![\alpha]\!]$. Thus L is $\preceq'$-closed.

From Lemma 1.4 (i), there exist a $\sim_{\mathcal{A}}$-class V and $i_1 < i_2 < \cdots$ such that $u(i_j, i_k) \in V$ for each $j < k$. Let U be a $\sim_{\mathcal{A}}$-class such that $u(1, i_1) \in U$. (We borrow the notation from Lemma 1.4 (i).) Since $\sim_{\mathcal{A}}$ is a congruence, $v(1, i_1) \in U$ and $v(i_j, i_k) \in V$ for each $j < k$. Thus $u_1 u_2 \cdots \in U.V^\omega$ implies $v_1 v_2 \cdots \in U.V^\omega$, and $\alpha \preceq \beta$. ∎

Definition 2.5 [1] For $u, v \in A^*$, we define $u \approx_L v$ if $w(w_1 u w_2)^\omega \in L \Leftrightarrow w(w_1 v w_2)^\omega \in L$ and $w_1 u w_2 w^\omega \in L \Leftrightarrow w_1 v w_2 w^\omega \in L$ for each $w, w_1, w_2 \in A^*$.

Proof of Theorem 2.2

Only-if part: Assume L is regular. Let $\mathcal{A}$ be a *Büchi* automaton such that $L = L(\mathcal{A})$. Since $\sim_{\mathcal{A}}$ is a finite congruence, $(A^*, \sim_{\mathcal{A}})$ is a monotone WQO. Define $\preceq$ as the transitive closure of $\preceq'$ (defined in Lemma 2.4), then $(A^\omega, \preceq)$ is a periodic extension of $(A^*, \sim_{\mathcal{A}})$ and $L(\mathcal{A})$ is $\preceq$-closed.

If part: Assume that L is $\preceq$-closed where $\preceq$ is a periodic extension of a monotone WQO $\leq$. First, we show that $\approx_L$ is a finite congruence. Assume that $\{u_i\}$ is an infinite set in A^* such that $u_i \not\approx_L u_j$ for $i \neq j$. Since $(A^*, \leq)$ is a WQO, there exists an infinite ascending subsequence $\{u_{k_i}\}$.

Let $F(u) = \{(v, v_1, v_2, w_1, w_2, w) \in A^* \times A^* \times A^* \times A^* \times A^* \times A^* \mid v(v_1 u v_2)^\omega \in L \ \wedge \ w_1 u w_2 w^\omega \in L\}$. Since $\preceq$ is a periodic extension of $\leq$ and L is $\preceq$-closed, each $F(u)$ is $\leq \times \leq \times \leq \times \leq \times \leq \times \leq$-closed and hence $F(u_{k_i}) \subseteq F(u_{k_j})$ for $i < j$. Since $u_{k_i} \not\approx_L u_{k_j}$ for $i \neq j$, $F(u_{k_i}) \neq F(u_{k_j})$, thus $F(u_{k_i}) \subset F(u_{k_j})$. Then there exists an infinite sequence in which each pair of different elements is incomparable. Since $\leq \times \leq \times \leq \times \leq \times \leq \times \leq$ is a WQO over $A^* \times A^* \times A^* \times A^* \times A^* \times A^*$, this is a contradiction.

Second, we show that $\approx_L$ saturates L. Assume that some $\approx_L$-classes U, V satisfy $U.V^\omega \cap L \neq \phi$ and $U.V^\omega \not\subseteq L$. From Lemma 1.4 (ii), we can assume that $U.V \subseteq U$ and $V.V \subseteq V$.

Let $\alpha \in U.V^\omega \cap L$ and $\beta \in U.V^\omega \setminus L$. Since $(A^\omega, \preceq)$ is a periodic extension, from Lemma 2.3 there exist $u, u' \in U$ and $v, v' \in V$ such that $\alpha = uv^\omega$ and $\beta = u'v'^\omega$. By definition of $\approx_L$, $uv^\omega \in L$ and $u'v'^\omega \notin L$ are contradictory. ∎

3 Second theorem

Definition 3.1 For a monotone QO $(A^*, \leq)$, a QO $(A^\omega, \preceq)$ is a *continuous extension* if the following conditions are satisfied.

(1) For each $u, v \in A^*$ and $\alpha, \beta \in A^\omega$, $u \leq v$ and $\alpha \preceq \beta$ imply $u\alpha \preceq v\beta$.

(2) Let $u_j, v_j \in A^*$ for each j and let $\alpha_i = v_1 \cdots v_{i-1} u_i \cdots$ for each i and $\alpha_\infty = v_1 v_2 \cdots$. For $\beta \in A^\omega$, if $u_i \leq v_i$ and $\alpha_i \preceq \beta$ for each i, then $\alpha_\infty \preceq \beta$, and if $u_i \geq v_i$ and $\alpha_i \succeq \beta$ for each i, then $\alpha_\infty \succeq \beta$.

Theorem 3.2 Let $L \subseteq A^\omega$. L is regular if and only if L is $\preceq$-closed under a WQO $(A^\omega, \preceq)$ that is a continuous extension of a monotone WQO $(A^*, \leq)$.

For the embedding $\leq$ over finite words, let $(A^*, \leq^\circ)$ be defined as $u \leq^\circ v$ if and only if $u \leq v$ and $elt(u) = elt(v)$, where $elt(u) = \{a_i \mid u = a_1 a_2 \cdots a_j\}$. Since the embedding $\leq$ over finite words is a WQO from Higman's lemma, $\leq^\circ$ is also a WQO. Then the embedding over A^ω is a continuous extension of $\leq^\circ$. Note that the embedding over A^ω is a continuous extension of the embedding $\leq$ over finite words. Actually, any continuous extension of the embedding $\leq$ over finite words is a trivial WQO (i.e., $A^\omega \times A^\omega$). For instance, given $\alpha, \beta \in A^\omega$. Let $\alpha(1, i)$ be the prefix of α of the length i and $\alpha_i = \alpha(1, i).\beta$ for each i. Since $\alpha(1, i) \geq \epsilon$, $\alpha_i \succeq \beta$ for each i. Thus, by definition of continuity, $\alpha_\infty = \alpha \succeq \beta$. Hence, for any $\alpha, \beta \in A^\omega$, we conclude $\alpha \succeq \beta$.

Definition 3.3 Let $u, v \in A^*$ and let $L \subseteq A^\omega$. We write

- $u \simeq^1_L v$ if and only if $\forall w \in A^*, \forall \alpha \in A^\omega.\ wu\alpha \in L \ \Leftrightarrow\ wv\alpha \in L$,
- $u \simeq^2_L v$ if and only if $\forall w \in A^*\ .\ wu^\omega \in L \ \Leftrightarrow\ wv^\omega \in L$, and
- $u \simeq_L v$ if and only if $u \simeq^1_L v$ and $u \simeq^2_L v$.

Proof of Theorem 3.2

Only-if part: Assume L is regular. Let $\mathcal{A}$ be a *Büchi* automaton such that $L = L(\mathcal{A})$. Since $\sim_\mathcal{A}$ is a finite congruence, $(A^*, \sim_\mathcal{A})$ is a monotone WQO. Define $\preceq$ as the transitive closure of $\preceq'$ (defined in Lemma 2.4), then $L(\mathcal{A})$ is $\preceq$-closed. Since $\preceq'$ is symmetric, $(A^\omega, \preceq)$ is a continuous extension of $(A^*, \sim_\mathcal{A})$ from Lemma 2.4 (ii). For the index n of $\sim_\mathcal{A}$, the number of $\preceq$-classes is bound by 2^{n^2}. Thus $\preceq$ is a WQO.

If part: First, we show that $\simeq_L$ is a finite congruence. Assume that $\{u_i\}$ is an infinite set in A^* such that $u_i \not\simeq_L u_j$ for $i \neq j$. Since $(A^*, \leq)$ is a WQO, there exists an infinite ascending subsequence $\{u_{k_i}\}$.

Let $F(u) \subseteq A^* \times A^\omega \times A^*$ be a set such that $(w, \alpha, v) \in F(u) \ \Leftrightarrow\ wu\alpha \in L \wedge vu^\omega \in L$. Then, each $F(u)$ is $\leq \times \preceq \times \leq$-closed and hence $F(u_{k_i}) \subseteq F(u_{k_j})$ for $i < j$. Since $u_{k_i} \not\simeq_L u_{k_j}$ for $i \neq j$, $F(u_{k_i}) \neq F(u_{k_j})$, thus $F(u_{k_i}) \subset F(u_{k_j})$. Then there exists an infinite sequence in which each pair of different elements is incomparable. Since $\leq \times \preceq \times \leq$ is a WQO over $A^* \times A^\omega \times A^*$, this is a contradiction.

Second, we show that $\simeq_L$ saturates L. Assume that some $\simeq_L$-classes U, V

satisfy $U.V^\omega \cap L \neq \phi$ and $U.V^\omega \not\subseteq L$. From Lemma 1.4 (2), we can assume that $V.V \subseteq V$.

Let $\alpha = uv_1v_2\cdots$ be a minimal element (wrt $\preceq$) in $U.V^\omega \cap L$, and let $\beta = u'v'_1v'_2\cdots \in U.V^\omega \setminus L$ such that $u, u' \in U$ and $v_i, v'_i \in V$. Let $\{\bar{v}_l\}$ be sets of minimal elements of V wrt $\leq$. Since $(V, \leq)$ is a WQO, $\{\bar{v}_l\}$ are finite.

Let $\alpha'(j, j+k) = v_j \cdots v_{j+k}$. Since $\bar{v}_l$ are finitely many, from (infinite) *Ramsey Theorem* there exist l and an ascending sequence $0 < j_1 < j_2 < \cdots$ such that $\alpha'(j_m, j_{m+1} - 1) \geq \bar{v}_l$ for any $m > 0$.

Let $\alpha_m = u\ \alpha'(1, j_1 - 1)\ \bar{v}_l^{m-1}\ \alpha'(j_m, j_{m+1} - 1)\cdots$. Obviously, $\alpha_m \preceq \alpha$ and $\alpha_m \in U.V^\omega \cap L$. Since α is minimal in $U.V^\omega \cap L$, $\alpha_m \succeq \alpha$. By definition of the continuous extension, $\alpha_\infty = u\ \alpha'(1, j_1 - 1)\ \bar{v}_l^\omega \succeq \alpha$. Thus since L is $\preceq$-closed, $\alpha_\infty \in U.V^\omega \cap L$.

Let $\beta'(j, j+k) = v'_j \cdots v'_{j+k}$. Since $\bar{v}_l$ are finitely many, from (infinite) *Ramsey Theorem* there exist l' and an ascending sequence $0 < j'_1 < j'_2 < \cdots$ such that $\beta'(j'_m, j'_{m+1} - 1) \geq \bar{v}_{l'}$ for any $m > 0$. Let $\beta_\infty = u'\ \beta'(1, j_1 - 1)\ \bar{v}_{l'}^\omega$. By definition of the continuous extension, $\beta_\infty \preceq \beta$. Since L is $\preceq$-closed, $\beta \notin L$ implies $\beta_\infty \notin L$. Thus $\bar{\beta} \in U.V^\omega \setminus L$.

Since $u \simeq_L^1 u'$ and $\bar{v}_j \simeq_L^2 \bar{v}_{j'}$ for each j, repeated applications of $\simeq_L^1$ and an application of $\simeq_L^2$ imply that $\alpha_\infty \in L \Leftrightarrow \beta_\infty \in L$. This contradicts $\alpha_\infty \in L$ and $\beta_\infty \notin L$. ■

Example 3.4 Either the periodic or continuous assumption cannot be dropped. Let $\beta = abaabaaabaaaab\cdots$ and let $L(\beta)$ be the set of ω-words that have a common suffix with β. For $\alpha \in A^\omega$, let $p_\beta(\alpha) = 1$ if $\alpha \in L(\beta)$ and let $p_\beta(\alpha) = 0$ if $\alpha \notin L(\beta)$. Define $\alpha \preceq \alpha' \Leftrightarrow p_\beta(\alpha) \leq p_\beta(\alpha')$. Then $\preceq$ is a WQO over ω-words and $L(\beta)$ is $\preceq$-closed, but $L(\beta)$ is not regular.

Acknowledgements

The author thanks Jean-Eric PIN for valuable comments at the previous presentation. This work is partially supported by PRESTO, Japan Science and Technology Corporation.

References

[1] A. Arnold. A syntactic congruence for rational ω-languages. *Theoretical Computer Science*, 39:333–335, 1985.

[2] A. Ehrenfeucht, D. Hausser, and G. Rozenberg. On regularity of context-free languages. *Theoretical Computer Science*, 27:311–332, 1983.

[3] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, chapter 4, pages 133–192. Elsevier Science Publishers, 1990.