

Combinatorics on Words formalized
Two Generated Word Monoids Intersection

Štěpán Holub
Štěpán Starosta

April 18, 2024

Funded by the Czech Science Foundation grant GAČR 20-20621S.

Contents

| | | |
|----------|---|----------|
| 1 | Binary Intersection Formalized | 2 |
| 1.1 | Blocks and intersection | 3 |
| 1.2 | Simple blocks | 5 |
| 1.3 | At least one block | 5 |
| 1.4 | Infinite case | 6 |
| 1.4.1 | Description of coincidence blocks | 6 |
| 1.5 | Description of the basis | 6 |
| 1.6 | Intersection | 7 |
| | References | 9 |

theory *Two-Generated-Word-Monoids-Intersection*

imports *Combinatorics-Words.Equations-Basic Combinatorics-Words.Binary-Code-Morphisms
Combinatorics-Words-Graph-Lemma.Glued-Codes*

begin

The characterization of intersection of binary languages formalized here is due to [1].

Chapter 1

Binary Intersection Formalized

locale *binary-codes-coincidence-two-generators* = *binary-codes-coincidence* +
assumes *two-coins*: $\exists r s r' s'. g r =_m h s \wedge g r' =_m h s' \wedge (r,s) \neq (r',s')$

begin

lemma *criticalE'*:

obtains $p q r1 s1 r2 s2$ **where**

$g p \cdot \alpha_g = h q \cdot \alpha_h$ **and**

$g (p \cdot r1) = h (q \cdot s1)$ **and**

$g (p \cdot r2) = h (q \cdot s2)$ **and**

$r1 \neq \varepsilon$ **and** $r2 \neq \varepsilon$ **and**

$hd r1 \neq hd r2$

$\langle proof \rangle$

lemma *alphas-suf*: $\alpha_h \leq_s \alpha_g$

$\langle proof \rangle$

lemma *c-def*: $c \cdot \alpha_h = \alpha_g$

$\langle proof \rangle$

lemma *marked-version-solution-conv*: $g_m r = h_m s \iff g r \cdot c = c \cdot h s$

$\langle proof \rangle$

lemma *criticalE*:

obtains $p q r1 s1 r2 s2$ **where**

$\alpha_g \cdot g_m p = \alpha_h \cdot h_m q$ **and**

$\bigwedge p' q'. \alpha_g \cdot g_m p' = \alpha_h \cdot h_m q' \implies p \leq_p p' \wedge q \leq_p q'$ **and**

$g_m (r1 \cdot p) = h_m (s1 \cdot q)$ **and**

$g_m (r2 \cdot p) = h_m (s2 \cdot q)$ **and**

$r1 \cdot p \neq \varepsilon$ **and** $r2 \cdot p \neq \varepsilon$ **and**

$hd (r1 \cdot p) \neq hd (r2 \cdot p)$

$\langle proof \rangle$

Defining the beginning block

definition *beginning-block* :: *binA list* * *binA list* **where**

$$\begin{aligned} \text{beginning-block} &= (\text{SOME pair}. \alpha_g \cdot g_m (\text{fst pair}) = \alpha_h \cdot h_m (\text{snd pair}) \wedge \\ &(\forall p' q'. \alpha_g \cdot g_m p' = \alpha_h \cdot h_m q' \longrightarrow (\text{fst pair}) \leq p p' \wedge (\text{snd pair}) \leq p q') \end{aligned}$$

definition *fst-beginning-block* (*p*) **where**

$$\text{fst-beginning-block} \equiv \text{fst beginning-block}$$

definition *snd-beginning-block* (*q*) **where**

$$\text{snd-beginning-block} \equiv \text{snd beginning-block}$$

lemma *begin-block*: $\alpha \cdot g_m p = h_m q$ **and**

$$\text{begin-block-min}: \alpha \cdot g_m p' = h_m q' \implies p \leq p p' \wedge q \leq p q'$$

<proof>

lemma *begin-block-conjug-conv*:

$$\text{assumes } r \cdot p = p \cdot r' \text{ and } s \cdot q = q \cdot s'$$

$$\text{shows } g r = h s \longleftrightarrow g_m r' = h_m s'$$

<proof>

lemma *solution-ext-conv*: $g r = h s \longleftrightarrow \alpha \cdot g_m (r \cdot p) = h_m (s \cdot q)$

<proof>

Both block exist

lemma *both-blocks*: *marked.blockP c*

<proof>

notation *marked.suc-fst* (**e**) **and**

$$\text{marked.suc-snd} (\mathbf{f})$$

lemma *sucs-eq*: $g_m (\mathbf{e} \tau) = h_m (\mathbf{f} \tau)$

<proof>

sublocale *marked*: *two-binary-marked-blocks* $g_m h_m$

<proof>

1.1 Blocks and intersection

Every solution has a block decomposition. However, not all block combinations yield a solution. This motivates the following definition.

definition *coin-block* **where** *coin-block* $\tau \equiv p \leq s p \cdot (\mathbf{e} \tau) \wedge q \leq s q \cdot (\mathbf{f} \tau)$

theorem *char-coincidence*:

$$g r = h s \longleftrightarrow (\exists \tau. \text{coin-block } \tau \wedge r = (p \cdot \mathbf{e} \tau)^{<-1} p \wedge s = (q \cdot \mathbf{f} \tau)^{<-1} q) \text{ (is } g$$

$$r = h s \longleftrightarrow ?Q)$$

<proof>

theorem *char-coincidence'*:

$g r = h s \iff (g_m (p^{-1} \triangleright (r \cdot p)) = h_m (q^{-1} \triangleright (s \cdot q)) \wedge p \leq p r \cdot p \wedge q \leq p s \cdot q)$
(is $g r = h s \iff ?Q$)
 <proof>

theorem coincidence-eq-blocks: $\mathfrak{C} g h = \{((p \cdot \mathfrak{e} \tau)^{<-1} p, (q \cdot \mathfrak{f} \tau)^{<-1} q) \mid \tau. \text{ coin-block } \tau\}$
 <proof>

lemma

minblock0: $g_m (\mathfrak{e} \mathfrak{a}) =_m h_m (\mathfrak{f} \mathfrak{a})$ **and**
minblock1: $g_m (\mathfrak{e} \mathfrak{b}) =_m h_m (\mathfrak{f} \mathfrak{b})$ **and**
hdblock0: $hd (\mathfrak{e} \mathfrak{a}) = \text{bina}$ **and**
hdblock1: $hd (\mathfrak{e} \mathfrak{b}) = \text{binb}$
 <proof>

definition \mathcal{T} where $\mathcal{T} \equiv \{\tau. \text{ coin-block } \tau\}$

lemma \mathcal{T} -def': $\tau \in \mathcal{T} \iff \text{ coin-block } \tau$
 <proof>

Properties of the set of coincidence blocks

lemma \mathcal{T} -closed: **assumes** *coin-block* τ_1 **and** *coin-block* τ_2
shows *coin-block* $(\tau_1 \cdot \tau_2)$
 <proof>

lemma emp-block: *coin-block* ε
 <proof>

lemma \mathcal{T} -hull: $\langle \mathcal{T} \rangle = \mathcal{T}$
 <proof>

lemma \mathcal{T} -pref: *coin-block* $\tau_1 \implies \text{ coin-block } (\tau_1 \cdot \tau_2) \implies \text{ coin-block } \tau_2$
 <proof>

Translation from blocks to the intersection

lemma translate-coin-blocks-to-intersection:
 $(h \circ (\lambda x. (q \cdot x)^{<-1} q) \circ \mathfrak{f}) \cdot \mathcal{T} = \text{range } g \cap \text{range } h$
 <proof>

lemma translation-blocks-inj:
inj-on $(h \circ (\lambda x. (q \cdot x)^{<-1} q) \circ \mathfrak{f}) \langle \mathcal{T} \rangle$
 <proof>

lemma translation-blocks-morph-on: *morphism-on* $(h \circ (\lambda x. (q \cdot x)^{<-1} q) \circ \mathfrak{f}) \mathcal{T}$
 <proof>

interpretation *morphism-on* $(h \circ (\lambda x. (q \cdot x)^{<-1} q) \circ \mathfrak{f}) \mathcal{T}$
 <proof>

theorem *inter-basis*: $\mathfrak{B} (\text{range } g \cap \text{range } h) = (h \circ (\lambda x. (q \cdot x)^{<-1} q) \circ f) \cdot (\mathfrak{B} \mathcal{T})$
 ⟨*proof*⟩

1.2 Simple blocks

If both letters are blocks, the situation is easy

theorem *simple-blocks*: **assumes** $\wedge a. \text{coin-block } [a]$ **shows** $\text{coin-block } \tau$
 ⟨*proof*⟩

theorem *simple-blocks-UNIV*: $(\wedge a. \text{coin-block } [a]) \implies \mathcal{T} = \text{UNIV}$
 ⟨*proof*⟩

theorem *simple-blocks-basis*: **assumes** $\wedge a. \text{coin-block } [a]$
shows $\mathfrak{B} \mathcal{T} = \{\mathfrak{a}, \mathfrak{b}\}$
 ⟨*proof*⟩

1.3 At least one block

At least one letter – the last one – is a block

lemma *last-letter-fst-suf*: **assumes** $\text{coin-block } (z \cdot [c])$
shows $p <_s \mathfrak{e} [c]$
 ⟨*proof*⟩

lemma *rich-block-suf-fst'*:
assumes $\text{coin-block } (z \cdot [1-c] \cdot [c]^{\text{@}} \text{Suc } i)$
shows $g_m.\text{bin-code-lcs} \cdot g_m p \leq_s g_m (\mathfrak{e} ([1-c] \cdot [c]^{\text{@}} \text{Suc } i))$
 ⟨*proof*⟩

lemma *rich-block-suf-fst*:
assumes $\text{coin-block } (z \cdot [1-c] \cdot [c]^{\text{@}} \text{Suc } i)$
shows $\alpha \cdot g_m (p) \leq_s g_m (\mathfrak{e} ([1-c] \cdot [c]^{\text{@}} \text{Suc } i))$
 ⟨*proof*⟩

lemma *rich-block-suf-snd'*:
assumes $\text{coin-block } (z \cdot [1-c] \cdot [c]^{\text{@}} \text{Suc } i)$
shows $\alpha_h \cdot h_m q \leq_s h_m (\mathfrak{f} ([1-c] \cdot [c]^{\text{@}} \text{Suc } i))$
 ⟨*proof*⟩

lemma *rich-block-suf-snd*:
assumes $\text{coin-block } (z \cdot [1-c] \cdot [c]^{\text{@}} \text{Suc } i)$
shows $q \leq_s \mathfrak{f} ([1-c] \cdot [c]^{\text{@}} \text{Suc } i)$
 ⟨*proof*⟩

lemma *last-letter-block*: **assumes** $\text{coin-block } (z \cdot [c])$
shows $\text{coin-block } [c]$

<proof>

end

1.4 Infinite case

locale *binary-codes-coincidence-infinite* = *binary-codes-coincidence-two-generators*
for *a1* +

assumes *non-block*: \neg *coin-block* [*a1*]

begin

1.4.1 Description of coincidence blocks

lemma *swap-coin-block*: *coin-block* [$1-a1$]

<proof>

definition *coincidence-exponent* (*t*) **where**

coincidence-exponent = (*LEAST* *x*. ($q \leq s$ $q \cdot f([a1] \cdot [1-a1]^{\textcircled{q}} \text{Suc } x)$))

lemma *q-nemp*: $q \neq \varepsilon$

<proof>

lemma *p-suf*: $p < s$ ε [$1-a1$]

<proof>

lemma *coin-exp*: *coin-block* ($[a1] \cdot [1-a1]^{\textcircled{q}} \text{Suc } t$) **and**

coin-exp-min: $j \leq t \implies \neg$ *coin-block* ($[a1] \cdot [1-a1]^{\textcircled{j}}$)

<proof>

lemma *exp-min*: \neg $q \leq s$ f [$1-a1$][ⓐ]*t*

<proof>

lemma *q-suf-conv*: $q \leq s$ f ($[a1] \cdot [1-a1]^{\textcircled{q}} \text{Suc } k$) \longleftrightarrow $t \leq k$

<proof>

lemma *coin-block-with-bad-letter*: **assumes** *a1* \in *set w*

shows *coin-block* *w* \longleftrightarrow [$1-a1$][ⓐ]*Suc t* $\leq s$ *w*

<proof>

1.5 Description of the basis

The infinite part of the basis

inductive-set \mathcal{W} :: *binA list set* **where**

$[a1] \cdot [1-a1]^{\textcircled{q}} \text{Suc } t \in \mathcal{W}$

$| \tau \in \mathcal{W} \implies i \leq t \implies [a1] \cdot [1-a1]^{\textcircled{i}} \cdot \tau \in \mathcal{W}$

lemma *W-nemp*: $x \in \mathcal{W} \implies x \neq \varepsilon$

<proof>

lemma *W-nemp'*: $x \in (\{[1 - a1]\} \cup \mathcal{W}) \implies x \neq \varepsilon$
<proof>

lemma *W-hd*: $x \in \mathcal{W} \implies hd\ x = a1$
<proof>

lemma *W-set*: $x \in \mathcal{W} \implies a1 \in set\ x$
<proof>

lemma *W-butlast-hd-tl*: $x \in \mathcal{W} \implies butlast\ x = [a1] \cdot butlast\ (tl\ x)$
<proof>

lemma *W-suf*: $x \in \mathcal{W} \implies [a1] \cdot [1-a1]^{\textcircled{a}} Suc\ t \leq_s x$
<proof>

lemma *W-fac*: $x \in \mathcal{W} \implies \neg [1-a1]^{\textcircled{a}} Suc\ t \leq_f butlast\ x$
<proof>

lemma *pref-code-W*: *pref-code* ($\{[1-a1]\} \cup \mathcal{W}$)
<proof>

lemma *W-coin-blocks*:
assumes $x \in \{[1 - a1]\} \cup \mathcal{W}$ **shows** $x \in \mathcal{T}$
<proof>

lemma *W-gen-T*: $\langle \{[1-a1]\} \cup \mathcal{W} \rangle = \mathcal{T}$
<proof>

lemma *W-explicit*: $\mathcal{W} = \{w \cdot [a1] \cdot [1-a1]^{\textcircled{a}} Suc\ t \mid w. w \in \langle \{[a1] \cdot [1-a1]^{\textcircled{a}} i \mid i. i \leq t \rangle \rangle$
<proof>

theorem *infinite-basis*: $\mathfrak{B}\ \mathcal{T} = (\{[1-a1]\} \cup \mathcal{W})$
<proof>

end

1.6 Intersection

lemma *bin-inter-coin-set-fst*: $\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle = ((bin-morph-of\ x\ y) \circ fst) \text{ ‘ } \mathfrak{C}$
(bin-morph-of x y) (bin-morph-of u v)
<proof>

lemma *bin-inter-coin-set-snd*: $\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle = ((bin-morph-of\ u\ v) \circ snd) \text{ ‘ } \mathfrak{C}$
(bin-morph-of x y) (bin-morph-of u v)
<proof>

theorem *bin-inter-basis*: **assumes** *binary-code x y and binary-code u v*
shows $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = ((\text{bin-morph-of } u \ v) \circ \text{snd}) \cdot \mathfrak{C}_m (\text{bin-morph-of } x \ y) (\text{bin-morph-of } u \ v)$
 $\langle \text{proof} \rangle$

theorem *binary-intersection-code*:
assumes *binary-code x y and binary-code u v*
shows *code* $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle)$
 $\langle \text{proof} \rangle$

theorem *binary-intersection*:
assumes *binary-code x y and binary-code u v*
obtains
 $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = \{\}$
 $|$
 β **where** $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = \{\beta\}$
 $|$
 $\beta \ \gamma$ **where** $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = \{\beta, \gamma\}$
 $|$
 $\beta \ \gamma \ \delta \ t$ **where** $\delta \neq \varepsilon$ **and** $\gamma \cdot \beta \neq \varepsilon$ **and** $hd \ \delta \neq hd \ (\gamma \cdot \beta)$
 $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = \{\beta \cdot \gamma\} \cup \{\beta \cdot (\gamma \cdot \beta)^{\otimes t} \cdot w \cdot \delta \cdot \gamma \mid w. w \in \langle \{\delta \cdot (\gamma \cdot \beta)^{\otimes i} \mid i. i \leq t\} \rangle\}$
 $|$
 $\beta \ \gamma \ \delta \ t \ q$ **where** $\delta \neq \varepsilon$ **and** $\gamma \cdot \beta \neq \varepsilon$ **and** $hd \ \delta \neq hd \ (\gamma \cdot \beta)$ **and**
 $1 \leq q \wedge q \leq t$ **and**
 $\mathfrak{B} (\langle \{x,y\} \rangle \cap \langle \{u,v\} \rangle) = \{\beta \cdot \gamma\} \cup \{\beta \cdot (\gamma \cdot \beta)^{\otimes t} \cdot w \cdot \delta^{<-1} (\beta \cdot (\gamma \cdot \beta)^{\otimes (t-q)}) \mid w. w \in \langle \{\delta \cdot (\gamma \cdot \beta)^{\otimes i} \mid i. i \leq q - 1\} \rangle\}$
 $|$
 $\langle \text{proof} \rangle$

end

References

- [1] J. Karhumäki. A note on intersections of free submonoids of a free monoid. *Semigroup forum*, 29:183–206, 1984.