

## PAPER

# Compactness of Finite Union of Regular Patterns and Regular Patterns without Adjacent Variables

Naoto TAKETA<sup>†</sup>, Nonmember, Tomoyuki UCHIDA<sup>†</sup>, Takayoshi SHOUDAI<sup>††</sup>, Satoshi MATSUMOTO<sup>†††</sup>, Yusuke SUZUKI<sup>†</sup>, and Tetsuhiro MIYAHARA<sup>†</sup>, Members

**SUMMARY** A regular pattern is a string consisting of constant symbols and distinct variable symbols. The language  $L(p)$  of a regular pattern  $p$  is the set of all constant strings obtained by replacing all variable symbols in the regular pattern  $p$  with constant strings.  $\mathcal{RP}^k$  denotes the class of all sets consisting at most  $k$  ( $k \geq 2$ ) regular patterns. Sato et al. (Proc. ALT'98, 1998) showed that the finite set  $S_2(P)$  of symbol strings is a characteristic set of  $L(P) = \bigcup_{p \in P} L(p)$ , where  $S_2(P)$  is obtained from  $P \in \mathcal{RP}^k$  by substituting variables with symbol strings of at most length 2. They also showed that  $\mathcal{RP}^k$  has compactness with respect to containment, if the number of constant symbols is greater than or equal to  $2k - 1$ . In this paper, we check their results and correct the error of the proof of their theorem. Further, we consider the set  $\mathcal{RP}_{NAV}$  of all non-adjacent regular patterns, which are regular patterns without adjacent variables, and show that the set  $S_2(P)$  obtained from a set  $P$  in the class  $\mathcal{RP}_{NAV}^k$  of at most  $k$  ( $k \geq 1$ ) non-adjacent regular patterns is a characteristic set of  $L(P)$ . Further we show that  $\mathcal{RP}_{NAV}^k$  has compactness with respect to containment if the number of constant symbols is greater than or equal to  $k + 2$ . Thus we show that we can design an efficient learning algorithm of a finite union of pattern languages of non-adjacent regular patterns with the number of constant symbols which is smaller than the case of regular patterns.

**key words:** Regular Pattern Language, Compactness w.r.t. Containment, Non-adjacent Regular Patterns Language

## 1. Introduction

A pattern is a string consisting of constant symbols and variable symbols [1], [2]. For example, we consider constant symbols  $a, b, c$  and variable symbols  $x, y$ , then  $axbxxy$  is a pattern.  $\mathcal{P}$  denotes the set of all patterns. For a pattern  $p \in \mathcal{P}$ , the pattern language generated by  $p$ , denoted by  $L(p)$ , or simply called a pattern language, is the set of all strings obtained by replacing all variable symbols with constant symbol strings, where the same variable symbol is replaced by the same constant string. For example the pattern language  $L(axbxxy)$  generated by the above pattern  $axbxxy$  denotes  $\{aubucw \mid u \text{ and } w \text{ are constant strings that are not } \varepsilon\}$ . A pattern where each variable symbol appears at most once is called a *regular pattern*. For example, a pattern  $axbxxy$  is not a regular pattern, but a pattern  $axbzcy$  with variable symbols  $x, y, z$  is a regular pattern.  $\mathcal{RP}$  denotes the set of all regular patterns.

If a pattern  $p \in \mathcal{P}$  is obtained from a pattern  $q \in \mathcal{P}$  by

Manuscript received January 1, 2015.

Manuscript revised January 1, 2015.

<sup>†</sup>Graduate School of Information Sciences, Hiroshima City University

<sup>††</sup>Department of Computer Science and Engineering, Fukuoka Institute of Technology

<sup>†††</sup>Faculty of Science, Tokai University

DOI: 10.1587/transinf.E0.D.1

replacing variable symbols in  $q$  with patterns, we say that  $q$  is a *generalization* of  $p$  and denote this by  $p \preceq q$ . For example, a pattern  $q = axz$  is a generalization of a pattern  $p = axbxxy$ , because  $p$  is obtained from  $q$  by replacing the variable  $z$  in  $q$  with a pattern  $bxy$ . So we write  $p \preceq q$ . For patterns  $p, q \in \mathcal{P}$ , it is obvious that  $p \preceq q$  implies  $L(p) \subseteq L(q)$ . But, the converse, that is, the statement that  $L(p) \subseteq L(q)$  implies  $p \preceq q$  does not always hold. With respect to this statement, Mukouchi [3] showed that if the number of constant symbols is greater than or equal to 3, for any regular pattern  $p, q \in \mathcal{RP}$ ,  $L(p) \subseteq L(q)$  implies  $p \preceq q$ .

We denote by  $\mathcal{RP}^+$  the class of all non-empty finite sets of regular patterns and by  $\mathcal{RP}^k$  the class of at most  $k$  ( $k \geq 2$ ) regular patterns. For a set of regular patterns  $P \in \mathcal{RP}^k$  we define  $L(P) = \bigcup_{p \in P} L(p)$  and consider the class  $\mathcal{RPL}^k$  of regular pattern languages of  $\mathcal{RP}^k$ , where  $\mathcal{RPL}^k = \{L(P) \mid P \in \mathcal{RP}^k\}$ . Let  $P, Q \in \mathcal{RP}^k$  and  $Q = \{q_1, \dots, q_k\}$ . We denote by  $P \sqsubseteq Q$  that for any regular pattern  $p \in P$  there exists a regular pattern  $q_i$  such that  $p \preceq q_i$  holds. From definition, it is obvious that  $P \sqsubseteq Q$  implies  $L(P) \subseteq L(Q)$ . Then Sato et al. [4] shows that if  $k \geq 3$  and the number of constant symbols is  $2k - 1$  then the finite set  $S_2(P)$  of constant symbols obtained from  $P \in \mathcal{RP}^k$  by substituting variable symbols with constant strings of at most 2 length is a characteristic set of  $L(P)$ , that is, for any regular pattern language  $L' \in \mathcal{RPL}^k$ ,  $S_2(P) \subseteq L'$  implies  $L(P) \subseteq L'$ . Thus they shows that the following three statements: (i)  $S_2(P) \subseteq L(Q)$ , (ii)  $P \sqsubseteq Q$  and (iii)  $L(P) \subseteq L(Q)$  are equivalent. But the Lemma 14 in [4], which is used in this results, contains an error. In this paper we correct this lemma and give a correct proof showing the equivalence of the three statements shown in [4]. Sato et al. [4] shows that  $\mathcal{RP}^k$  has compactness with respect to containment if the number of constant symbols is greater than or equal to  $2k - 1$ . On the contrary to this result, we show that the set  $S_2(P)$  obtained from a set  $P$  in the class  $\mathcal{RP}_{NAV}^k$  of at most  $k$  ( $k \geq 1$ ) regular patterns having non-adjacent variables is a characteristic set of  $L(P)$ . Further, we show that if the number of constant symbols is greater than or equal to  $k + 2$  then  $\mathcal{RP}_{NAV}^k$  has compactness with respect to containment. In Table 1 we summarize the all results in this paper.

**Table 1** The conditions of the number of constant symbols with respect to the compactness of inclusion

$k$	2	$\geq 3$
$\mathcal{RP}^k$	$\geq 4$	$\geq 2k - 1$
$\mathcal{RP}_{NAV}^k$		$\geq k + 2$

Mukouchi [5] examined the decision problem of determining whether a containment relation exists between the languages generated by two given patterns. The inductive inference of formal languages—specifically, pattern languages [2] and unions of pattern languages [6], [7] from positive data has been extensively investigated. Arimura et al. [8] introduced a formal framework for the efficient generalization of unions of pattern languages, presenting a polynomial-time algorithm to identify the minimal set of patterns whose union encompasses a given set of positive examples. In a subsequent study, Arimura et al. [9] proposed the concept of strong compactness of containment for unions of regular pattern languages. Day et al. [10] established that pattern languages are, in general, not closed under standard language operations such as union, intersection, and complement. Matsumoto et al. [11] developed an efficient query learning algorithm for regular pattern languages that requires only a single positive example and a linear number of membership queries. More recently, Takeda et al. [12] proposed a query learning algorithm that utilizes a deep learning model trained on a set of strings as an oracle, enabling the learned features to be visualized as regular patterns. Subsequent research extended the study of regular patterns to Elementary Formal Systems (EFS) [13], thereby broadening the theoretical foundation of pattern languages. This extension inspired further work on tree patterns [14], [15] for generating tree languages, as well as on the development of Formal Graph Systems [16]. These advancements have facilitated the formalization and efficient learning of increasingly complex structured data beyond strings, fostering applications in domains such as grammatical inference and graph-based learning.

This paper is organized as follows. In Sect.2 as preparations, we give definitions of pattern languages, regular pattern languages and compactness, and then introduce the results of Sato et al.[4]. In Sect.3, we show that  $S_2(P)$  is a characteristic set of  $L(P)$  in  $\mathcal{RPL}^k$  and  $\mathcal{RP}^k$  has compactness with respect to containment. In Sect.4, we propose regular patterns having non-adjacent variables, show that  $S_2(P)$  obtained from a set  $P$  in  $\mathcal{RP}_{NAV}^k$  is a characteristic set of  $L(P)$ , and  $\mathcal{RP}_{NAV}^k$  has compactness with respect to containment.

## 2. Preliminaries

### 2.1 Basic definitions and notations

Let  $\Sigma$  be a non-empty finite set of constant symbols. Let  $X$  be an infinite set of variable symbols such that  $\Sigma \cap X = \emptyset$  holds. Then, a *string* on  $\Sigma \cup X$  is a sequence of symbols in  $\Sigma \cup X$ . Particularly, the string having no symbol is called the *empty string* and is denoted by  $\varepsilon$ . We denote by  $(\Sigma \cup X)^*$  the set of all strings on  $\Sigma \cup X$  and by  $(\Sigma \cup X)^+$  the set of all strings on  $\Sigma \cup X$  except  $\varepsilon$ , i.e.,  $(\Sigma \cup X)^+ = (\Sigma \cup X)^* \setminus \{\varepsilon\}$ .

A *pattern* on  $\Sigma \cup X$  is a string in  $(\Sigma \cup X)^*$ . Note that the empty string  $\varepsilon$  is a pattern on  $\Sigma \cup X$ . A pattern  $p$  is said to be *regular* if each variable symbol appears at most once in  $p$ .

The length of  $p$ , denote by  $|p|$ , is the number of symbols in  $p$ . Note that  $|\varepsilon| = 0$  holds. The set of all patterns and regular patterns on  $\Sigma \cup X$  are denoted by  $\mathcal{P}$  and  $\mathcal{RP}$ , respectively. For a set  $S$ , we denote by  $\#S$  the number of elements in  $S$ . Let  $p, q$  be strings. If  $p$  and  $q$  are equal as strings, we denote it by  $p = q$ . We denote by  $p \cdot q$  the string obtained from  $p$  and  $q$  by concatenating  $q$  after  $p$ . Note that for a string  $p$  and the empty string  $\varepsilon$ ,  $p \cdot \varepsilon = \varepsilon \cdot p = p$ .

A substitution  $\theta$  is a mapping from  $(\Sigma \cup X)^*$  to  $(\Sigma \cup X)^*$  such that (1)  $\theta$  is a homomorphism with respect to string concatenation, i.e.,  $\theta(p \cdot q) = \theta(p) \cdot \theta(q)$  holds for patterns  $p$  and  $q$ , (2)  $\theta(\varepsilon) = \varepsilon$  holds, (3) for each constant symbol  $a \in \Sigma$ ,  $\theta(a) = a$  holds, and (4) for each variable symbol  $x \in X$ ,  $|\theta(x)| \geq 1$  holds. Let  $x_1, \dots, x_n$  be variable symbols and  $p_1, \dots, p_n$  non-empty patterns. The notation  $\{x_1 := p_1, \dots, x_n := p_n\}$  denotes a substitution that replaces each variable symbol  $x_i$  with a non-empty pattern  $p_i$  for each  $i \in \{1, \dots, n\}$ . For a pattern  $p$  and a substitution  $\theta = \{x_1 := p_1, \dots, x_n := p_n\}$ , we denote by  $p\theta$  a new pattern obtained from  $p$  by replacing variable symbols  $x_1, \dots, x_n$  in  $p$  with patterns  $p_1, \dots, p_n$  according to  $\theta$ , respectively.

For a pattern  $p$  and  $q$ , the pattern  $q$  is a *generalization* of  $p$ , or  $p$  is an *instance* of  $q$ , denoted by  $p \preceq q$ , if there exists a substitution  $\theta$  such that  $p = q\theta$  holds. If  $p \preceq q$  and  $p \succeq q$  hold, we denote it by  $p \equiv q$ . The notation  $p \equiv q$  means that  $p$  and  $q$  are equal as strings except for variable symbols. For a pattern  $p$ , the *pattern language* of  $p$ , denoted by  $L(p)$ , is the set  $\{w \in \Sigma^* \mid w \preceq p\}$ . For patterns  $p$  and  $q$ , it is clear that  $L(p) = L(q)$  if  $p \equiv q$ , and  $L(p) \subseteq L(q)$  if  $p \preceq q$ . Note that  $L(\varepsilon) = \{\varepsilon\}$ . In particular, if  $p$  is a regular pattern, we say that  $L(p)$  is a *regular pattern language*. The set of all pattern languages and regular patterns languages are denoted by  $\mathcal{PL}$  and  $\mathcal{RPL}$ , respectively.

**Lemma 1** (Mukouchi(Theorem 6.1, [3])): Suppose  $\#\Sigma \geq 3$ . Let  $p$  and  $q$  be regular patterns. Then  $p \preceq q$  if and only if  $L(p) \subseteq L(q)$ .

Next, we consider unions of pattern languages. The class of all non-empty finite subsets of  $\mathcal{P}$  is denoted by  $\mathcal{P}^+$ , i.e.,  $\mathcal{P}^+ = \{P \subseteq \mathcal{P} \mid 0 < \#P < \infty\}$ . For a positive integer  $k$  i.e.,  $k > 0$ , the class of non-empty sets consisting of at most  $k$  patterns, i.e.,  $\mathcal{P}^k = \{P \subseteq \mathcal{P} \mid 0 < \#P \leq k\}$ . For a set  $P$  of patterns, the pattern language of  $P$ , denoted by  $L(P)$ , is the set  $\bigcup_{p \in P} L(p)$ . We denote by  $\mathcal{PL}^k$  the class of unions of at most  $k$  pattern languages, i.e.,  $\mathcal{PL}^k = \{L(P) \mid P \in \mathcal{P}^k\}$ . In a similar way, we also define  $\mathcal{RP}^+$ ,  $\mathcal{RP}^k$  and  $\mathcal{RPL}^k$ . For  $P, Q$  in  $\mathcal{P}^+$ , the notation  $P \sqsubseteq Q$  means that for any  $p \in P$  there is a pattern  $q \in Q$  such that  $p \preceq q$  holds. It is clear that  $P \sqsubseteq Q$  implies  $L(P) \subseteq L(Q)$ . However, the converse is not valid in general.

### 2.2 Characteristic sets

**Definition 1:** Let  $C$  be a class of languages,  $L$  a language in  $C$  and  $S$  a non-empty finite subset of  $L$ . We say that  $S$  is a *characteristic set* of  $L$  within  $C$  if for any  $L' \in C$ ,  $S \subseteq L'$  implies  $L \subseteq L'$ .

Let  $n$  be a positive integer and  $p$  a regular pattern. We denote by  $S_n(p)$  the set of all strings in  $\Sigma^*$  obtained by replacing all variable symbols in  $p$  with strings in  $\Sigma^+$  of length at most  $n$ . Moreover, for a positive integer  $n$  and a set  $P \in \mathcal{RP}^+$ , let  $S_n(P) = \bigcup_{p \in P} S_n(p)$ . It is clear that  $S_n(P) \subseteq S_{n+1}(P) \subseteq L(P)$  for any positive integer  $n$ .

**Theorem 1** (Sato et al.(Theorem 8, [4])): Let  $k$  be a positive integer and  $P \in \mathcal{RP}^k$ . Then, there exists a positive integer  $n$  such that  $S_n(P)$  is a characteristic set of  $L(P)$  within  $\mathcal{RPL}^k$ .

**Theorem 2** (Sato et al.(Lemma 9, [4])): Let  $p, q, p_1, p_2, q_1, q_2, q_3$  be regular patterns in  $\mathcal{RP}$  and  $x$  a variable symbol such that  $p = p_1xp_2$  and  $q = q_1q_2q_3$  hold. Then  $p \preceq q$  if the following three conditions (i), (ii) and (iii) are holds:

- (i)  $p_1 \preceq q_1q_2$ ,
- (ii)  $p_2 \preceq q_2q_3$ ,
- (iii)  $q_2$  contains at least one variable symbol.

**Lemma 2** (Sato et al.(Lemma 10, [4])): Suppose  $\#\Sigma \geq 3$ . Let  $p_1, p_2, q$  be regular patterns in  $\mathcal{RP}$  and  $x$  a variable symbol. Let  $a, b$  and  $c$  be mutually distinct constant symbols in  $\Sigma$ . If  $p_1ap_2 \preceq q$ ,  $p_1bp_2 \preceq q$  and  $p_1cp_2 \preceq q$ , then  $p_1xp_2 \preceq q$  holds.

**Lemma 3** (Sato et al.(Lemma 13, [4])): Suppose  $\#\Sigma \geq 3$ . Let  $p_1, p_2, q_1, q_2$  be regular patterns in  $\mathcal{RP}$  and  $x$  a variable symbol. Let  $a, b$  be constant symbols in  $\Sigma$  with  $a \neq b$  and  $w$  a string on  $\Sigma^*$ . Let  $p = p_1AwxwBp_2$  and  $q = q_1AwBq_2$  be regular patterns that satisfy the following three conditions:

- (i)  $p_1Aw \preceq q_1$ ,
- (ii)  $wBp_2 \preceq q_2$ ,
- (iii)  $(A, B) \in \{(a, b), (b, a)\}$ .

Then, we have that  $p\{x := a\} \preceq q$  and  $p\{x := b\} \preceq q$  hold but  $p \not\preceq q$ .

From Lemma 2, the following theorem holds.

**Theorem 3** (Sato et al.(Theorem 10, [4])): Let  $k$  be a positive integer. Suppose  $\#\Sigma \geq 2k + 1$ . For  $P \in \mathcal{RP}^+$  and  $Q \in \mathcal{RP}^k$ , the following (i), (ii) and (iii) are equivalent:

- (i)  $S_1(P) \subseteq L(Q)$ ,
- (ii)  $P \sqsubseteq Q$ ,
- (iii)  $L(P) \subseteq L(Q)$ .

The following Example 1 in ?? shows that Theorem 3 does not hold if  $\#\Sigma \leq 2k$  holds.

**Example 1:** Let  $k$  be a positive integer and  $\Sigma = \{a_1, \dots, a_k, b_1, \dots, b_k\}$ . Let  $w_1, \dots, w_k$  be regular patterns in  $\mathcal{RP}$  such that  $w_k = \varepsilon$  and for  $i = 1, 2, \dots, k-1$ ,  $w_i = w_{i+1}b_{i+1}a_{i+1}w_{i+1}$  hold. Let  $p, q_1, \dots, q_k$  be regular patterns in  $\mathcal{RP}$  such that  $p = x_1a_1w_1xw_1b_1x_2$  and for  $i = 1, 2, \dots, k$ ,  $q_i = x_1a_iw_ib_ix_2$  hold. Let  $Q$  be a set  $\{q_1, \dots, q_k\}$  in  $\mathcal{RP}^k$ . For  $i = 1$ , we have  $p\{x := a_1\} = (x_1a_1w_1)a_1(w_1b_1x_2) = q_1\{x_1 := x_1a_1w_1\} \preceq q_1$ . For  $i \geq 2$ , from the definition of  $w_i$ , we easily see that  $w_1 = (w_i b_i)w^{(i)} = w'(i)(a_i w_i)$  for some strings  $w^{(i)}$  and  $w'^{(i)}$ . Then, for each  $i \geq 2$ ,

$$p\{x := a_1\} = x_1a_1w_1a_iw_iw_1b_ix_2 = x_1a_1w_2b_2a_2w_2a_2w_2b_2a_2w_2b_2x_2 \dots$$

$$\begin{aligned} p\{x := a_i\} &= (x_1a_1w_1)a_i(w_1b_1x_2) \\ &= (x_1a_1w_1)a_i(w_i b_i w^{(i)})b_1x_2 \\ &= (x_1a_1w_1)(a_i w_i b_i)(w^{(i)} b_1 x_2) \\ &= q_i\{x_1 := x_1a_1w_1, x_2 := w^{(i)} b_1 x_2\} \\ &\preceq q_i, \\ p\{x := b_i\} &= (x_1a_1w_1)b_i(w_1b_1x_2) \\ &= x_1a_1(w'^{(i)} a_i w_i)b_i(w_1b_1x_2) \\ &= (x_1a_1w'^{(i)})a_i w_i b_i(w_1b_1x_2) \\ &= q_i\{x_1 := x_1a_1w'^{(i)}, x_2 := w_1b_1x_2\} \\ &\preceq q_i. \end{aligned}$$

Hence,  $S_1(p) \subseteq L(Q)$  holds. However, from  $p \not\preceq q_i$ ,  $L(p) \not\subseteq L(q_i)$  holds for each  $i = 1, 2, \dots, k$ .

From Theorem 3, we have the following corollary.

**Corollary 1** (Sato et al.(Corollary 12, [4])): Suppose  $\#\Sigma \geq 3$ . For two regular patterns  $p$  and  $q$ , the following (i), (ii) and (iii) are equivalent:

- (i)  $S_1(p) \subseteq L(q)$ ,
- (ii)  $p \preceq q$ ,
- (iii)  $L(p) \subseteq L(q)$ .

## 2.3 Basic word equations

**Proposition 1:** Let  $w$  be a string in  $\Sigma^*$  and  $a, b$  constant symbols in  $\Sigma$ . If

$$wa = bw \tag{1}$$

holds, then  $a = b$  holds.

**Proof.** Since it is trivial, we omit the proof.  $\square$

**Proposition 2:** Let  $w$  be a string in  $\Sigma^*$  and  $a, b, c, d$  constant symbols in  $\Sigma$ . If

$$wda = bcw \tag{2}$$

holds, then  $(b, c) \in \{(a, d), (d, a)\}$  holds.

**Proof.** We will prove this proposition by induction on the length of  $w$ , i.e.,  $|w|$ .

- $|w| = 0, 1, 2, 3$ : it is straightforward to observe that  $(b, c) \in \{(a, d), (d, a)\}$  holds.
- $|w| \geq 4$ : We assume that for any string  $u$  with  $0 \leq |u| < n$ , if  $uda = bcu$  holds,  $(b, c) \in \{(a, d), (d, a)\}$  holds. Since the string  $w$  has a prefix  $bc$  and a suffix  $da$ , there exists a string  $u$  with  $|u| = |w| - 4 < |w|$  such that  $w = bcuda$  holds. Since  $wda = bcw$ , we have  $bcudada = bcbcuda$ , and then  $uda = bcu$ . Thus, from the assumption, we get  $(b, c) \in \{(a, d), (d, a)\}$ .

From the above, we conclude that if  $wda = bcw$  holds, then  $(b, c) \in \{(a, d), (d, a)\}$  holds.  $\square$

The conclusion from Proposition 2 shows that  $(a, d) \in \{(b, c), (c, b)\}$ . Therefore, if the equation  $daw = wbc$  holds, we arrive at the same conclusion.

**Proposition 3:** Let  $w, w'$  be strings of constant symbols in  $\Sigma$  and  $a, b, c, d$  constant symbols in  $\Sigma$ . If

$$wdaw' = w'bcw \quad (3)$$

holds, then  $(b, c) \in \{(a, d), (d, a)\}$  holds.

**Proof.** We will prove this proposition by an induction on  $|w| + |w'|$ . Without loss of generality, we assume that  $|w| \geq |w'|$  because, if  $|w| < |w'|$ , we arrive at the same conclusion that  $(a, d) \in \{(b, c), (c, b)\}$  holds.

- $|w| \geq 0$  and  $|w'| = 0$ : Eq. (3) reduces to  $wda = bcw$ . By Proposition 2,  $(b, c) \in \{(a, d), (d, a)\}$  holds.

We assume that for constant strings  $u$  and  $u'$  with  $|u| + |u'| < |w| + |w'|$ , if  $uda u' = u'bcu$  holds, then  $(b, c) \in \{(a, d), (d, a)\}$  holds. We divide the relations between  $|w|$  and  $|w'|$  into the following four cases:

- $0 < |w'| \leq |w| \leq |w'| + 1$ : When either  $|w| = |w'|$  or  $|w| = |w'| + 1$ , Eq. (3) is illustrated in Figs. 1 and 2, respectively. If  $|w| = |w'|$ ,  $(b, c) = (d, a)$  holds. If  $|w| = |w'| + 1$ ,  $d = c$  and  $w = w'b = aw'$  hold. From Proposition 1, we deduce that  $b = a$ . Therefore,  $(b, c) \in \{(a, d), (d, a)\}$  holds.
- $|w'| + 2 \leq |w| \leq 2|w'| - 1$ : In Eq. 3, since  $|wdaw'| = |w'bcw| = |w| + |w'| + 2$ , a suffix of  $w$  overlaps with a prefix of  $w$ , as illustrated in Fig. 3. That is, there exists a constant string  $u$  of length  $2|w| - (|w| + |w'| + 2) = |w| - |w'| - 2$  such that  $u$  is both a prefix and a suffix of  $w$ . Since  $uda$  has a length of  $|w| - |w'|$ , it is also a prefix of  $w$ . Similarly,  $bcu$  is a suffix of  $w$ . Because  $|w| - (|uda| + |bcu|) = 2|w| - |w'| \geq 1$ , there exists a constant string  $u'$  of length  $2|w'| - |w|$  such that  $w = uda u'bcu$  holds. Since  $w'$  is a suffix of  $w$  and  $|u'bcu| = (2|w'| - |w|) + 2 + (|w| - |w'| - 2) = |w'|$ , we have  $w' = u'bcu$ . Similarly,  $w' = uda u'$ . Thus, we derive the equation  $u'bcu = uda u'$ . Since  $|u| = |w| - |w'| - 2 \leq |w| - 3 < |w|$  and  $|u'| = 2|w'| - |w| < |w'|$ , i.e.,  $|u| + |u'| < |w| + |w'|$ , the induction hypothesis on  $|u| + |u'|$  implies that  $(b, c) \in \{(a, d), (d, a)\}$  holds.
- $2|w'| \leq |w| \leq 2|w'| + 3$ : When  $|w| = 2|w'|$ , it is straightforward to observe that  $w = w'w'$ . Therefore,  $w'da = bcw'$  holds, as illustrated in Fig. 4. From Proposition 2,  $(b, c) \in \{(a, d), (d, a)\}$  holds. When  $|w| = 2|w'| + i$  ( $i = 1, 2, 3$ ), Eq. (3) is depicted in Figs. 5, 6, and 7, respectively. When  $|w| = 2|w'| + 2$ , it is clear that  $(b, c) = (d, a)$ . When  $|w| = 2|w'| + 1$  and  $|w| = 2|w'| + 3$ , Proposition 1 implies that  $(b, c) = (a, d)$  holds.
- $2|w'| + 4 \leq |w|$ : Since the strings  $w'bc$  and  $adw'$  are a prefix and a suffix of  $w$ , respectively, and  $|w'bc| + |adw'| = 2|w'| + 4$ , there exists a string  $u$  with  $|u| \geq 0$  such that  $w = w'bcudaw'$  holds. From Eq. (3),  $w'bcudaw'daw' = w'bcw'bcudaw'$ , i.e.,  $udaw' = w'bcu$  holds, as illustrated in Fig. 8. Let  $u' = w'$ . Since  $|u| + |u'| = |w| - (2|w'| + 4) + |w'| < |w| + |w'|$ , the induction hypothesis on  $|u| + |u'|$  implies that  $(b, c) \in \{(a, d), (d, a)\}$  holds.

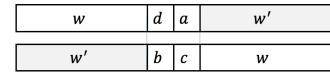


Fig. 1 Case  $|w| = |w'|$  in Proposition 3

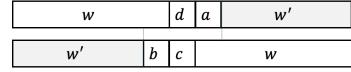


Fig. 2 Case  $|w| = |w'| + 1$  in Proposition 3

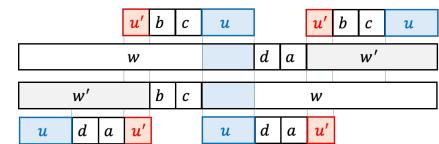


Fig. 3 Case  $|w'| + 2 \leq |w| \leq 2|w'| - 1$  in Proposition 3

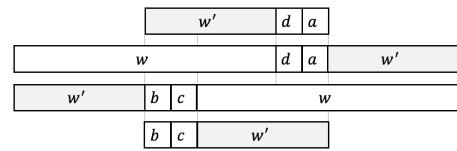


Fig. 4 Case  $|w| = 2|w'|$  in Proposition 3

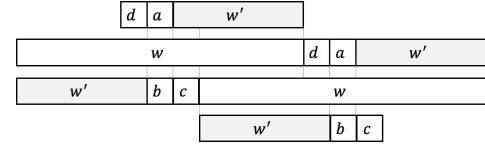


Fig. 5 Case  $|w| = 2|w'| + 1$  in Proposition 3

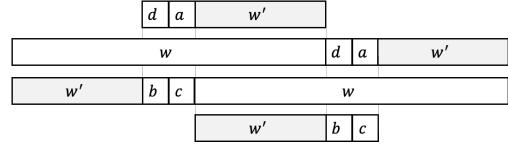


Fig. 6 Case  $|w| = 2|w'| + 2$  in Proposition 3

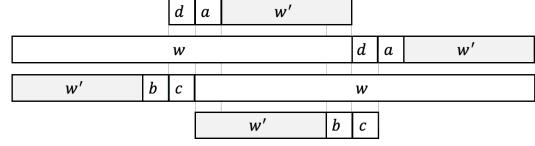


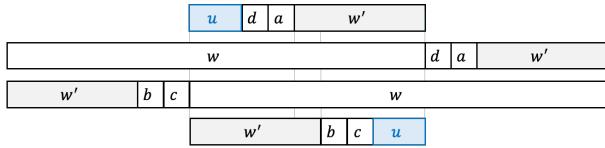
Fig. 7 Case  $|w| = 2|w'| + 3$  in Proposition 3

From the above, we conclude that if  $wdaw' = w'bcw$ , then  $(b, c) \in \{(a, d), (d, a)\}$  holds.  $\square$

### 3. Compactness for Sets of Regular Patterns

#### 3.1 Compactness

In this section, we define the compactness of sets of regular patterns, formally. Then, if  $\#\Sigma \geq 2k - 1$  holds, we show that

Fig. 8 Case 2  $|w'| + 4 \leq |w|$  in Proposition 3

$\mathcal{RP}^k$  has compactness with respect to the containment.

**Definition 2:** Let  $C$  be a subset of  $\mathcal{RP}^+$ . For any regular pattern  $p \in \mathcal{RP}$  and any set  $Q \in C$ , the set  $C$  said to have **compactness with respect to containment** if there exists a regular pattern  $q \in Q$  such that  $L(p) \subseteq L(q)$  holds if  $L(p) \subseteq L(Q)$  holds.

Let  $D \subset \mathcal{RP}$  with  $\#D = 2$  or  $3$ , and let  $p, q$  be regular patterns in  $\mathcal{RP}$ . In the following subsections (Subsecs. 3.2–3.5), we provide the conditions on  $D$  under which the implication holds: if  $p\{x := r\} \preceq q$  for all  $r \in D$ , then  $p\{x := xy\} \preceq q$ . It is obvious if the variable symbol  $x$  does not appear in  $p$ . Therefore, in the following lemmas and propositions, let  $p = p_1xp_2$ , where  $p_i \in \mathcal{RP}$  ( $i = 1, 2$ ) and  $x$  is a variable symbol.

First of all, we consider the correspondence from  $r \in D$  to some string in  $q$  when  $p\{x := r\} \preceq q$  holds. The symbols in  $D$  correspond to either a variable or a constant symbol in  $q$ . If  $D$  has a constant string  $ab$  of length 2 for  $a, b \in \Sigma$ , there are three possible strings in  $q$  that correspond to  $ab$  in  $p\{x := ab\}$  as follows: For  $y_1 \in X$ ,

- (a)  $ab$ ,
- (b)  $ay_1$ ,
- (c)  $y_1b$ .

If there exists (b)  $ay_1$  in  $q$  that corresponds to  $ab$ , i.e., there exist  $q_1$  and  $q_2 \in \mathcal{RP}$  such that

- (1)  $p_1abp_2 \preceq q_1ay_1q_2$ ,
- (2)  $p_1 \preceq q_1$ , and
- (3) either  $p_2 \preceq q_2$  or  $p_2 \preceq y'_1q_2$  for  $y'_1 \in X$ .

Let  $D' = (D \setminus \{ab\}) \cup \{ay\}$ . It is straightforward to see that  $p\{x := ay\} = p_1ayp_2 \preceq q_1ay_1q_2$  holds. Thus,  $p\{x := r\} \preceq q$  for all  $r \in D'$  holds. Let  $D'' = (D \setminus \{ab\}) \cup \{yb\}$ . By a similar discussion, if there exists (c)  $y_1b$  in  $q$  that corresponds to  $ab$ ,  $p\{x := r\} \preceq q$  for all  $r \in D''$  holds. Therefore, in this paper, we make the following definition on  $D$ :

**Definition 3:** Let  $p, q \in \mathcal{RP}$ . Let  $D \subset \mathcal{RP}$  such that for all  $r \in D$ ,  $|r| = 2$  and  $p\{x := r\} \preceq q$  holds. Then, if for any  $ab \in D$  ( $a, b \in \Sigma$ ),  $p\{x := ay\} \not\preceq q$  and  $p\{x := yb\} \not\preceq q$  hold for any  $y \in X$  that does not appear in  $q$ , the set  $D$  is said to be **maximally generalized on**  $(p, q)$ .

### 3.2 $D = \{ay, by\}$ and $D = \{ya, yb\}$

**Lemma 4** (Sato et al.[4]): Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$  and let  $p, q$  be regular patterns on  $\Sigma \cup X$ . Let  $D$  be the following set of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol that does not appear in  $p$  and  $q$ :

Note that  $p \neq q$

$\# \Sigma \geq 2$   
 $\# \Sigma \geq 3$   
 $\# \Sigma \geq 4$   
 $\# \Sigma \geq 5$   
 $\# \Sigma \geq 6$   
 $\# \Sigma \geq 7$   
 $\# \Sigma \geq 8$   
 $\# \Sigma \geq 9$   
 $\# \Sigma \geq 10$   
 $\# \Sigma \geq 11$   
 $\# \Sigma \geq 12$   
 $\# \Sigma \geq 13$   
 $\# \Sigma \geq 14$   
 $\# \Sigma \geq 15$   
 $\# \Sigma \geq 16$   
 $\# \Sigma \geq 17$   
 $\# \Sigma \geq 18$   
 $\# \Sigma \geq 19$   
 $\# \Sigma \geq 20$   
 $\# \Sigma \geq 21$   
 $\# \Sigma \geq 22$   
 $\# \Sigma \geq 23$   
 $\# \Sigma \geq 24$   
 $\# \Sigma \geq 25$   
 $\# \Sigma \geq 26$   
 $\# \Sigma \geq 27$   
 $\# \Sigma \geq 28$   
 $\# \Sigma \geq 29$   
 $\# \Sigma \geq 30$   
 $\# \Sigma \geq 31$   
 $\# \Sigma \geq 32$   
 $\# \Sigma \geq 33$   
 $\# \Sigma \geq 34$   
 $\# \Sigma \geq 35$   
 $\# \Sigma \geq 36$   
 $\# \Sigma \geq 37$   
 $\# \Sigma \geq 38$   
 $\# \Sigma \geq 39$   
 $\# \Sigma \geq 40$   
 $\# \Sigma \geq 41$   
 $\# \Sigma \geq 42$   
 $\# \Sigma \geq 43$   
 $\# \Sigma \geq 44$   
 $\# \Sigma \geq 45$   
 $\# \Sigma \geq 46$   
 $\# \Sigma \geq 47$   
 $\# \Sigma \geq 48$   
 $\# \Sigma \geq 49$   
 $\# \Sigma \geq 50$   
 $\# \Sigma \geq 51$   
 $\# \Sigma \geq 52$   
 $\# \Sigma \geq 53$   
 $\# \Sigma \geq 54$   
 $\# \Sigma \geq 55$   
 $\# \Sigma \geq 56$   
 $\# \Sigma \geq 57$   
 $\# \Sigma \geq 58$   
 $\# \Sigma \geq 59$   
 $\# \Sigma \geq 60$   
 $\# \Sigma \geq 61$   
 $\# \Sigma \geq 62$   
 $\# \Sigma \geq 63$   
 $\# \Sigma \geq 64$   
 $\# \Sigma \geq 65$   
 $\# \Sigma \geq 66$   
 $\# \Sigma \geq 67$   
 $\# \Sigma \geq 68$   
 $\# \Sigma \geq 69$   
 $\# \Sigma \geq 70$   
 $\# \Sigma \geq 71$   
 $\# \Sigma \geq 72$   
 $\# \Sigma \geq 73$   
 $\# \Sigma \geq 74$   
 $\# \Sigma \geq 75$   
 $\# \Sigma \geq 76$   
 $\# \Sigma \geq 77$   
 $\# \Sigma \geq 78$   
 $\# \Sigma \geq 79$   
 $\# \Sigma \geq 80$   
 $\# \Sigma \geq 81$   
 $\# \Sigma \geq 82$   
 $\# \Sigma \geq 83$   
 $\# \Sigma \geq 84$   
 $\# \Sigma \geq 85$   
 $\# \Sigma \geq 86$   
 $\# \Sigma \geq 87$   
 $\# \Sigma \geq 88$   
 $\# \Sigma \geq 89$   
 $\# \Sigma \geq 90$   
 $\# \Sigma \geq 91$   
 $\# \Sigma \geq 92$   
 $\# \Sigma \geq 93$   
 $\# \Sigma \geq 94$   
 $\# \Sigma \geq 95$   
 $\# \Sigma \geq 96$   
 $\# \Sigma \geq 97$   
 $\# \Sigma \geq 98$   
 $\# \Sigma \geq 99$   
 $\# \Sigma \geq 100$   
 $\# \Sigma \geq 101$   
 $\# \Sigma \geq 102$   
 $\# \Sigma \geq 103$   
 $\# \Sigma \geq 104$   
 $\# \Sigma \geq 105$   
 $\# \Sigma \geq 106$   
 $\# \Sigma \geq 107$   
 $\# \Sigma \geq 108$   
 $\# \Sigma \geq 109$   
 $\# \Sigma \geq 110$   
 $\# \Sigma \geq 111$   
 $\# \Sigma \geq 112$   
 $\# \Sigma \geq 113$   
 $\# \Sigma \geq 114$   
 $\# \Sigma \geq 115$   
 $\# \Sigma \geq 116$   
 $\# \Sigma \geq 117$   
 $\# \Sigma \geq 118$   
 $\# \Sigma \geq 119$   
 $\# \Sigma \geq 120$   
 $\# \Sigma \geq 121$   
 $\# \Sigma \geq 122$   
 $\# \Sigma \geq 123$   
 $\# \Sigma \geq 124$   
 $\# \Sigma \geq 125$   
 $\# \Sigma \geq 126$   
 $\# \Sigma \geq 127$   
 $\# \Sigma \geq 128$   
 $\# \Sigma \geq 129$   
 $\# \Sigma \geq 130$   
 $\# \Sigma \geq 131$   
 $\# \Sigma \geq 132$   
 $\# \Sigma \geq 133$   
 $\# \Sigma \geq 134$   
 $\# \Sigma \geq 135$   
 $\# \Sigma \geq 136$   
 $\# \Sigma \geq 137$   
 $\# \Sigma \geq 138$   
 $\# \Sigma \geq 139$   
 $\# \Sigma \geq 140$   
 $\# \Sigma \geq 141$   
 $\# \Sigma \geq 142$   
 $\# \Sigma \geq 143$   
 $\# \Sigma \geq 144$   
 $\# \Sigma \geq 145$   
 $\# \Sigma \geq 146$   
 $\# \Sigma \geq 147$   
 $\# \Sigma \geq 148$   
 $\# \Sigma \geq 149$   
 $\# \Sigma \geq 150$   
 $\# \Sigma \geq 151$   
 $\# \Sigma \geq 152$   
 $\# \Sigma \geq 153$   
 $\# \Sigma \geq 154$   
 $\# \Sigma \geq 155$   
 $\# \Sigma \geq 156$   
 $\# \Sigma \geq 157$   
 $\# \Sigma \geq 158$   
 $\# \Sigma \geq 159$   
 $\# \Sigma \geq 160$   
 $\# \Sigma \geq 161$   
 $\# \Sigma \geq 162$   
 $\# \Sigma \geq 163$   
 $\# \Sigma \geq 164$   
 $\# \Sigma \geq 165$   
 $\# \Sigma \geq 166$   
 $\# \Sigma \geq 167$   
 $\# \Sigma \geq 168$   
 $\# \Sigma \geq 169$   
 $\# \Sigma \geq 170$   
 $\# \Sigma \geq 171$   
 $\# \Sigma \geq 172$   
 $\# \Sigma \geq 173$   
 $\# \Sigma \geq 174$   
 $\# \Sigma \geq 175$   
 $\# \Sigma \geq 176$   
 $\# \Sigma \geq 177$   
 $\# \Sigma \geq 178$   
 $\# \Sigma \geq 179$   
 $\# \Sigma \geq 180$   
 $\# \Sigma \geq 181$   
 $\# \Sigma \geq 182$   
 $\# \Sigma \geq 183$   
 $\# \Sigma \geq 184$   
 $\# \Sigma \geq 185$   
 $\# \Sigma \geq 186$   
 $\# \Sigma \geq 187$   
 $\# \Sigma \geq 188$   
 $\# \Sigma \geq 189$   
 $\# \Sigma \geq 190$   
 $\# \Sigma \geq 191$   
 $\# \Sigma \geq 192$   
 $\# \Sigma \geq 193$   
 $\# \Sigma \geq 194$   
 $\# \Sigma \geq 195$   
 $\# \Sigma \geq 196$   
 $\# \Sigma \geq 197$   
 $\# \Sigma \geq 198$   
 $\# \Sigma \geq 199$   
 $\# \Sigma \geq 200$   
 $\# \Sigma \geq 201$   
 $\# \Sigma \geq 202$   
 $\# \Sigma \geq 203$   
 $\# \Sigma \geq 204$   
 $\# \Sigma \geq 205$   
 $\# \Sigma \geq 206$   
 $\# \Sigma \geq 207$   
 $\# \Sigma \geq 208$   
 $\# \Sigma \geq 209$   
 $\# \Sigma \geq 210$   
 $\# \Sigma \geq 211$   
 $\# \Sigma \geq 212$   
 $\# \Sigma \geq 213$   
 $\# \Sigma \geq 214$   
 $\# \Sigma \geq 215$   
 $\# \Sigma \geq 216$   
 $\# \Sigma \geq 217$   
 $\# \Sigma \geq 218$   
 $\# \Sigma \geq 219$   
 $\# \Sigma \geq 220$   
 $\# \Sigma \geq 221$   
 $\# \Sigma \geq 222$   
 $\# \Sigma \geq 223$   
 $\# \Sigma \geq 224$   
 $\# \Sigma \geq 225$   
 $\# \Sigma \geq 226$   
 $\# \Sigma \geq 227$   
 $\# \Sigma \geq 228$   
 $\# \Sigma \geq 229$   
 $\# \Sigma \geq 230$   
 $\# \Sigma \geq 231$   
 $\# \Sigma \geq 232$   
 $\# \Sigma \geq 233$   
 $\# \Sigma \geq 234$   
 $\# \Sigma \geq 235$   
 $\# \Sigma \geq 236$   
 $\# \Sigma \geq 237$   
 $\# \Sigma \geq 238$   
 $\# \Sigma \geq 239$   
 $\# \Sigma \geq 240$   
 $\# \Sigma \geq 241$   
 $\# \Sigma \geq 242$   
 $\# \Sigma \geq 243$   
 $\# \Sigma \geq 244$   
 $\# \Sigma \geq 245$   
 $\# \Sigma \geq 246$   
 $\# \Sigma \geq 247$   
 $\# \Sigma \geq 248$   
 $\# \Sigma \geq 249$   
 $\# \Sigma \geq 250$   
 $\# \Sigma \geq 251$   
 $\# \Sigma \geq 252$   
 $\# \Sigma \geq 253$   
 $\# \Sigma \geq 254$   
 $\# \Sigma \geq 255$   
 $\# \Sigma \geq 256$   
 $\# \Sigma \geq 257$   
 $\# \Sigma \geq 258$   
 $\# \Sigma \geq 259$   
 $\# \Sigma \geq 260$   
 $\# \Sigma \geq 261$   
 $\# \Sigma \geq 262$   
 $\# \Sigma \geq 263$   
 $\# \Sigma \geq 264$   
 $\# \Sigma \geq 265$   
 $\# \Sigma \geq 266$   
 $\# \Sigma \geq 267$   
 $\# \Sigma \geq 268$   
 $\# \Sigma \geq 269$   
 $\# \Sigma \geq 270$   
 $\# \Sigma \geq 271$   
 $\# \Sigma \geq 272$   
 $\# \Sigma \geq 273$   
 $\# \Sigma \geq 274$   
 $\# \Sigma \geq 275$   
 $\# \Sigma \geq 276$   
 $\# \Sigma \geq 277$   
 $\# \Sigma \geq 278$   
 $\# \Sigma \geq 279$   
 $\# \Sigma \geq 280$   
 $\# \Sigma \geq 281$   
 $\# \Sigma \geq 282$   
 $\# \Sigma \geq 283$   
 $\# \Sigma \geq 284$   
 $\# \Sigma \geq 285$   
 $\# \Sigma \geq 286$   
 $\# \Sigma \geq 287$   
 $\# \Sigma \geq 288$   
 $\# \Sigma \geq 289$   
 $\# \Sigma \geq 290$   
 $\# \Sigma \geq 291$   
 $\# \Sigma \geq 292$   
 $\# \Sigma \geq 293$   
 $\# \Sigma \geq 294$   
 $\# \Sigma \geq 295$   
 $\# \Sigma \geq 296$   
 $\# \Sigma \geq 297$   
 $\# \Sigma \geq 298$   
 $\# \Sigma \geq 299$   
 $\# \Sigma \geq 300$   
 $\# \Sigma \geq 301$   
 $\# \Sigma \geq 302$   
 $\# \Sigma \geq 303$   
 $\# \Sigma \geq 304$   
 $\# \Sigma \geq 305$   
 $\# \Sigma \geq 306$   
 $\# \Sigma \geq 307$   
 $\# \Sigma \geq 308$   
 $\# \Sigma \geq 309$   
 $\# \Sigma \geq 310$   
 $\# \Sigma \geq 311$   
 $\# \Sigma \geq 312$   
 $\# \Sigma \geq 313$   
 $\# \Sigma \geq 314$   
 $\# \Sigma \geq 315$   
 $\# \Sigma \geq 316$   
 $\# \Sigma \geq 317$   
 $\# \Sigma \geq 318$   
 $\# \Sigma \geq 319$   
 $\# \Sigma \geq 320$   
 $\# \Sigma \geq 321$   
 $\# \Sigma \geq 322$   
 $\# \Sigma \geq 323$   
 $\# \Sigma \geq 324$   
 $\# \Sigma \geq 325$   
 $\# \Sigma \geq 326$   
 $\# \Sigma \geq 327$   
 $\# \Sigma \geq 328$   
 $\# \Sigma \geq 329$   
 $\# \Sigma \geq 330$   
 $\# \Sigma \geq 331$   
 $\# \Sigma \geq 332$   
 $\# \Sigma \geq 333$   
 $\# \Sigma \geq 334$   
 $\# \Sigma \geq 335$   
 $\# \Sigma \geq 336$   
 $\# \Sigma \geq 337$   
 $\# \Sigma \geq 338$   
 $\# \Sigma \geq 339$   
 $\# \Sigma \geq 340$   
 $\# \Sigma \geq 341$   
 $\# \Sigma \geq 342$   
 $\# \Sigma \geq 343$   
 $\# \Sigma \geq 344$   
 $\# \Sigma \geq 345$   
 $\# \Sigma \geq 346$   
 $\# \Sigma \geq 347$   
 $\# \Sigma \geq 348$   
 $\# \Sigma \geq 349$   
 $\# \Sigma \geq 350$   
 $\# \Sigma \geq 351$   
 $\# \Sigma \geq 352$   
 $\# \Sigma \geq 353$   
 $\# \Sigma \geq 354$   
 $\# \Sigma \geq 355$   
 $\# \Sigma \geq 356$   
 $\# \Sigma \geq 357$   
 $\# \Sigma \geq 358$   
 $\# \Sigma \geq 359$   
 $\# \Sigma \geq 360$   
 $\# \Sigma \geq 361$   
 $\# \Sigma \geq 362$   
 $\# \Sigma \geq 363$   
 $\# \Sigma \geq 364$   
 $\# \Sigma \geq 365$   
 $\# \Sigma \geq 366$   
 $\# \Sigma \geq 367$   
 $\# \Sigma \geq 368$   
 $\# \Sigma \geq 369$   
 $\# \Sigma \geq 370$   
 $\# \Sigma \geq 371$   
 $\# \Sigma \geq 372$   
 $\# \Sigma \geq 373$   
 $\# \Sigma \geq 374$   
 $\# \Sigma \geq 375$   
 $\# \Sigma \geq 376$   
 $\# \Sigma \geq 377$   
 $\# \Sigma \geq 378$   
 $\# \Sigma \geq 379$   
 $\# \Sigma \geq 380$   
 $\# \Sigma \geq 381$   
 $\# \Sigma \geq 382$   
 $\# \Sigma \geq 383$   
 $\# \Sigma \geq 384$   
 $\# \Sigma \geq 385$   
 $\# \Sigma \geq 386$   
 $\# \Sigma \geq 387$   
 $\# \Sigma \geq 388$   
 $\# \Sigma \geq 389$   
 $\# \Sigma \geq 390$   
 $\# \Sigma \geq 391$   
 $\# \Sigma \geq 392$   
 $\# \Sigma \geq 393$   
 $\# \Sigma \geq 394$   
 $\# \Sigma \geq 395$   
 $\# \Sigma \geq 396$   
 $\# \Sigma \geq 397$   
 $\# \Sigma \geq 398$   
 $\# \Sigma \geq 399$   
 $\# \Sigma \geq 400$   
 $\# \Sigma \geq 401$   
 $\# \Sigma \geq 402$   
 $\# \Sigma \geq 403$   
 $\# \Sigma \geq 404$   
 $\# \Sigma \geq 405$   
 $\# \Sigma \geq 406$   
 $\# \Sigma \geq 407$   
 $\# \Sigma \geq 408$   
 $\# \Sigma \geq 409$   
 $\# \Sigma \geq 410$   
 $\# \Sigma \geq 411$   
 $\# \Sigma \geq 412$   
 $\# \Sigma \geq 413$   
 $\# \Sigma \geq 414$   
 $\# \Sigma \geq 415$   
 $\# \Sigma \geq 416$   
 $\# \Sigma \geq 417$   
 $\# \Sigma \geq 418$   
 $\# \Sigma \geq 419$   
 $\# \Sigma \geq 420$   
 $\# \Sigma \geq 421$   
 $\# \Sigma \geq 422$   
 $\# \Sigma \geq 423$   
 $\# \Sigma \geq 424$   
 $\# \Sigma \geq 425$   
 $\# \Sigma \geq 426$   
 $\# \Sigma \geq 427$   
 $\# \Sigma \geq 428$   
 $\# \Sigma \geq 429$   
 $\# \Sigma \geq 430$   
 $\# \Sigma \geq 431$   
 $\# \Sigma \geq 432$   
 $\# \Sigma \geq 433$   
 $\# \Sigma \geq 434$   
 $\# \Sigma \geq 435$   
 $\# \Sigma \geq 436$   
 $\# \Sigma \geq 437$   
 $\# \Sigma \geq 438$   
 $\# \Sigma \geq 439$   
 $\# \Sigma \geq 440$   
 $\# \Sigma \geq 441$   
 $\# \Sigma \geq 442$   
 $\# \Sigma \geq 443$   
 $\# \Sigma \geq 444$   
 $\# \Sigma \geq 445$   
 $\# \Sigma \geq 446$   
 $\# \Sigma \geq 447$   
 $\# \Sigma \geq 448$   
 $\# \Sigma \geq 449$   
 $\# \Sigma \geq 450$   
 $\# \Sigma \geq 451$   
 $\# \Sigma \geq 452$   
 $\# \Sigma \geq 453$   
 $\# \Sigma \geq 454$   
 $\# \Sigma \geq 455$   
 $\# \Sigma \geq 456$   
 $\# \Sigma \geq 457$   
 $\# \Sigma \geq 458$   
 $\# \Sigma \geq 459$   
 $\# \Sigma \geq 460$   
 $\# \Sigma \geq 461$   
 $\# \Sigma \geq 462$   
 $\# \Sigma \geq 463$   
 $\# \Sigma \geq 464$   
 $\# \Sigma \geq 465$   
 $\# \Sigma \geq 466$   
 $\# \Sigma \geq 467$   
 $\# \Sigma \geq 468$   
 $\# \Sigma \geq 469$   
 $\# \Sigma \geq 470$   
 $\# \Sigma \geq 471$   
 $\# \Sigma \geq 472$   
 $\# \Sigma \geq 473$   
 $\# \Sigma \geq 474$   
 $\# \Sigma \geq 475$   
 $\# \Sigma \geq 476$   
 $\# \Sigma \geq 477$   
 $\# \Sigma \geq 478$   
 $\# \Sigma \geq 479$   
 $\# \Sigma \geq 480$   
 $\# \Sigma \geq 481$   
 $\# \Sigma \geq 482$   
 $\# \Sigma \geq 483$   
 $\# \Sigma \geq 484$   
 $\# \Sigma \geq 485$   
 $\# \Sigma \geq 486$   
 $\# \Sigma \geq 487$   
 $\# \Sigma \geq 488$   
 $\# \Sigma \geq 489$   
 $\# \Sigma \geq 490$   
 $\# \Sigma \geq 491$   
 $\# \Sigma \geq 492$   
 $\# \Sigma \geq 493$   
 $\# \Sigma \geq 494$   
 $\# \Sigma \geq 495$   
 $\# \Sigma \geq 496$   
 $\# \Sigma \geq 497$   
 $\# \Sigma \geq 498$   
 $\# \Sigma \geq 499$   
 $\# \Sigma \geq 500$   
 $\# \Sigma \geq 501$   
 $\# \Sigma \geq 502$   
 $\# \Sigma \geq 503$   
 $\# \Sigma \geq 504$   
 $\# \Sigma \geq 505$   
 $\# \Sigma \geq 506$   
 $\# \Sigma \geq 507$   
 $\# \Sigma \geq 508$   
 $\# \Sigma \geq 509$   
 $\# \Sigma \geq 510$   
 $\# \Sigma \geq 511$   
 $\# \Sigma \geq 512$   
 $\# \Sigma \geq 513$   
 $\# \Sigma \geq 514$   
 $\# \Sigma \geq 515$   
 $\# \Sigma \geq 516$   
 $\# \Sigma \geq 517$   
 $\# \Sigma \geq 518$   
 $\# \Sigma \geq 519$   
 $\# \Sigma \geq 520$   
 $\# \Sigma \geq 521$   
 $\# \Sigma \geq 522$   
 $\# \Sigma \geq 523$   
 $\# \Sigma \geq 524$   
 $\# \Sigma \geq 525$   
 $\# \Sigma \geq 526$   
 $\# \Sigma \geq 527$   
 $\# \Sigma \geq 528$ <

regular patterns on  $\Sigma \cup X$ . Let  $D$  be the following set of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol in  $X$  that does not appear in  $p$  and  $q$ :

$$D = \{ya, bc, dy\} \quad (b \notin \{a, d\} \text{ and } c \notin \{a, d\})$$

Then, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$ .

**Proof.** We assume that  $p\{x := xy\} \not\preceq q$  in order to derive a contradiction. Since  $D$  is maximally generalized on  $(p, q)$ , the regular pattern  $q$  can be expressed in one of the following forms: Let  $y_1, y_2$  be distinct variable symbols in  $X$  and  $q_1, q_2, w, w'$  be either the empty string or a regular pattern on  $\Sigma \cup X$ .

- (5-1)  $q = q_1AwBw'Cq_2$ ,  
where  $\{A, B, C\} = \{y_1a, bc, dy_2\}$ ,
- (5-2)  $q = q_1AwBq_2$ ,  
where  $\{A, B\} = \{dy_1a, bc\}$ ,
- (5-3)  $q = q_1AwBq_2$ ,  
where  $\{A, B\} = \{y_1ay_2, bc\}$  ( $a = d$ ).

(5-1) Case of  $q = q_1AwBw'Cq_2$ , where  $\{A, B, C\} = \{y_1a, bc, dy_2\}$ : At first, we prove the following three claims:

*Claim 1.*  $B \notin \{y_1a, dy_2\}$ .

*Proof of Claim 1.* Suppose that  $(A, B, C) = (dy_2, y_1a, bc)$ . The following conditions must be satisfied: For  $y'_1, y'_2 \in X$ ,

- |   |  |
|---|--|
| (1) $p_1 \preceq q_1$ ,                                       | (1') $p_2 \preceq wy_1aw'bcq_2$ or<br>$p_2 \preceq y'_2wy_1aw'bcq_2$ , |
| (2) $p_1 \preceq q_1dy_2w$ or<br>$p_1 \preceq q_1dy_2wy'_1$ , | (2') $p_2 \preceq w'bcq_2$ ,   |
| (3) $p_1 \preceq q_1dy_2wy_1aw'$ ,                            | (3') $p_2 \preceq q_2$ .   |

When  $p_2 \preceq wy_1aw'bcq_2$  in (1') holds, let  $q'_1 = q_1dy_2$ ,  $q'_2 = wy_1aw'$ ,  $q'_3 = bcq_2$ . Since  $p_1 \preceq q_1dy_2wy_1aw'$  holds from (3), both  $p_1 \preceq q'_1q'_2$  and  $p_2 \preceq q'_2q'_3$  hold, and  $q'_2$  contains a variable symbol. When  $p_2 \preceq y'_2wy_1aw'bcq_2$  in (1') holds, let  $q'_1 = q_1d$ ,  $q'_2 = y_2wy_1aw'$ ,  $q'_3 = bcq_2$ . Since  $p_1 \preceq q_1dy_2wy_1aw'$  holds from (3), both  $p_1 \preceq q'_1q'_2$  and  $p_2 \preceq q'_2q'_3$  hold, and  $q'_2$  contains a variable symbol. In both cases, by Theorem 2,  $p \preceq q$  holds. This contradicts the assumption that  $p\{x := xy\} \not\preceq q$ .

Similarly, we can show that any case where  $(A, B, C) = (y_1a, dy_2, bc)$ ,  $(bc, y_1a, dy_2)$ , or  $(bc, dy_2, y_1a)$  also contradicts the assumption. Therefore, we have  $B \notin \{y_1a, dy_2\}$ . (End of Proof of Claim 1)

*Claim 2.*  $(A, B, C) = (y_1a, bc, dy_2)$

*Proof of Claim 2.* From *Claim 1*, we have  $B = bc$ . Suppose that  $(A, B, C) = (dy_2, bc, y_1a)$ , i.e.,  $q = q_1dy_2wbcw'y_1aq_2$  holds. Then, the following conditions must be satisfied: For  $y'_1, y'_2 \in X$ ,

- |                                   |  |
|-----------------------------------|--|
| (1) $p_1 \preceq q_1$ ,           | (1') $p_2 \preceq wbcw'y_1aq_2$ or<br>$p_2 \preceq y'_2wbcw'y_1aq_2$ , |
| (2) $p_1 \preceq q_1dy_2w$ ,      | (2') $p_2 \preceq w'y_1aq_2$ ,   |
| (3) $p_1 \preceq q_1dy_2wbcw'$ or | (3') $p_2 \preceq q_2$ .   |

$$p_1 \preceq q_1dy_2wbcw'y_1'$$

From  $p_1 \preceq q_1dy_2w$  in (2),  $p_1$  is expressed as  $p'_1p''_1$  for some  $p'_1$  and  $p''_1$ , where  $p'_1 \preceq q_1d$  and  $p''_1 \preceq y_2w$ . When  $p_2 \preceq wbcw'y_1aq_2$  in (1'), we have  $p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1dp''_1wbcw'y_1aq_2 = q\{y_2 := p''_1x\}$ . Thus,  $p\{x := xy\} \preceq q\{y_2 := p''_1xy\}$  holds. This contradicts the assumption that  $p\{x := xy\} \not\preceq q$ . When  $p_2 \preceq y'_2wbcw'y_1aq_2$  in (1'), we similarly have  $p = p_1xp_2 = p'_1p''_1xp_2 \preceq q_1dp''_1xy'_2wbcw'y_1aq_2 = q\{y_2 := p''_1xy'_2\}$ . Thus,  $p\{x := xy\} \preceq q\{y_2 := p''_1xy'_2\}$  holds. This also contradicts the assumption. Therefore, we conclude that  $(A, B, C) = (y_1a, bc, dy_2)$ . (End of Proof of Claim 2)

From *Claim 2*, The regular pattern  $q$  is expressed as  $q_1y_1awbcw'dy_2q_2$ , where  $b \notin \{a, d\}$  and  $c \notin \{a, d\}$ . If  $p\{x := xy\} \not\preceq q$  holds, the following conditions must be satisfied: For  $y'_1, y'_2 \in X$ ,

- |  |   |
|--|---|
| (1) $p_1 \preceq q_1$ or $p_1 \preceq q_1y'_1$ , | (1') $p_2 \preceq wbcw'dy_2q_2$ ,                 |
| (2) $p_1 \preceq q_1y_1aw$ ,                     | (2') $p_2 \preceq w'dy_2q_2$ ,                    |
| (3) $p_1 \preceq q_1y_1awbcw'$ ,                 | (3') $p_2 \preceq q_2$ or $p_2 \preceq y'_2q_2$ . |

*Claim 3.*  $w$  and  $w'$  contain no variable symbols.

*Proof of Claim 3.* Let  $q'_1 = q_1y_1a$ ,  $q'_2 = wbcw'$ , and  $q'_3 = dy_2q_2$ . From (1') and (3),  $p_1 \preceq q'_1q'_2$  and  $p_2 \preceq q'_2q'_3$ . If  $q'_2$  contains a variable symbol, then by Theorem 2,  $p \preceq q$  holds. This contradicts the assumption. Therefore,  $w$  and  $w'$  contain no variable symbols. (End of Proof of Claim 3)

From *Claim 3*,  $w$  and  $w'$  are strings consisting of symbols in  $\Sigma$ . From (1') and (2'),  $wbcw'd$  and  $w'd$  are prefixes of  $p_2$ , and from (2) and (3),  $awbcw'$  and  $aw$  are suffixes of  $p_1$ . It implies a contradiction in the following inductive way:

- $|w| = |w'|$ : Directly,  $b = d$  and  $a = c$  hold.
- $|w| = |w'| + 1$ : Also,  $a = b$  holds.
- $|w| = |w'| + 2$ : Since  $awbcw'$  and  $aw$  are suffixes of  $p_1$ , and  $|w| \geq 2$ ,  $a$  is a suffix of  $w$ . From (1') and (2'), we have  $w = w'da$ . Furthermore, since  $awbcw'$  and  $aw$  are suffixes of  $p_1$ , it follows that  $w = bcw'$ . Thus,  $w'da = bcw'$  holds. From Proposition 2,  $(b, c) \in \{(a, d), (d, a)\}$  holds. Therefore, these cases contradict the conditions  $b \notin \{a, d\}$  and  $c \notin \{a, d\}$ .
- $|w| \geq |w'| + 3$ : From (2) and (3), there exists a string  $w''$  of length  $|w| - |w'| - 2$  such that  $w = w''bcw'$  holds. Moreover, from (2) and (3), since  $|aw| < |wbcw'|$  and  $aw = aw''bcw'$ , it follows that  $aw''$  is a suffix of  $w$ . On the other hand, from (1') and (2'),  $w'd$  is a prefix of  $w$ . Since  $|w'd| + |aw''| = |w'| + |w''| + 2 = |w|$ , it follows that  $w = w'daw''$  (Fig. 10). Therefore,  $w'daw'' = w''bcw'$  holds. From Proposition 3,  $(b, c) \in \{(a, d), (d, a)\}$  holds. This contradicts the conditions  $b \notin \{a, d\}$  and  $c \notin \{a, d\}$ .

From the above, we conclude that all cases of (5-1) contradict the assertion that  $p\{x := xy\} \not\preceq q$  and the conditions  $b \notin \{a, d\}$  and  $c \notin \{a, d\}$ .

(5-2) Case of  $q = q_1AwBq_2$ , where  $\{A, B\} = \{dy_1a, bc\}$ :

$|w| < |w'|$  同様に示すと並べて並ぶ  
書くべきか?

$\{a, b, c, d\}$   $\Sigma \cup X$   $b=a, b \neq d, a \neq c$   
 $= \{a, c, d\}$   $\Sigma \cap X$   $b \neq c$

$a$	$w$			(2)	
$a$	$w$	$b$	$c$	$w'$	(3)
	$a w'' b c w'$				
	$w$	$b$	$c$	$w'$	(1')
	$w b c w' d$			(2')	

Fig. 10 Case (5-1) in Lemma 5: Relation of strings  $w$ ,  $w'$ , and  $w''$

We suppose that  $(A, B) = (dy_1a, bc)$ , i.e.,  $q = q_1dy_1awbcq_2$  holds. Then, the following conditions must be satisfied for  $y'_1 \in X$ :

- |  |  |
|--|--|
| (1) $p_1 \preceq q_1$ ,                          | (1') $p_2 \preceq awbcq_2$ or<br>$p_2 \preceq y'_1awbcq_2$ , |
| (2) $p_1 \preceq q_1d$ or $p_1 \preceq q_1dy'_1$ | (2') $p_2 \preceq wbcq_2$ ,                                  |
| (3) $p_1 \preceq q_1dy_1aw$ ,                    | (3') $p_2 \preceq q_2$ .                                     |

From  $p_1 \preceq q_1dy_1aw$  in (3),  $p_1$  can be expressed as  $p'_1p''_1$  for some  $p'_1$  and  $p''_1$ , where  $p'_1 \preceq q_1d$  and  $p''_1 \preceq y_1aw$ . When  $p_2 \preceq awbcq_2$  in (1'), we have

$$p = p'_1p''_1xp_2 \preceq q_1dp''_1xawbcq_2 = q\{y_1 := p''_1x\}.$$

Thus,  $p\{x := xy\} \preceq q\{y_1 := p''_1xy\}$  holds. This contradicts the assumption. When  $p_2 \preceq y'_1awbcq_2$  in (1'), we similarly have

$$p = p'_1p''_1xp_2 \preceq q_1dp''_1xy'_1wbcq_2 = q\{y_1 := p''_1xy'_1\}.$$

Thus,  $p\{x := xy\} \preceq q\{y_1 := p''_1xyy'_1\}$  holds. This contradicts the assumption that  $p\{x := xy\} \not\preceq q$ . Similarly, we can show that the case  $(A, B) = (bc, dy_1a)$  also contradicts the assumption.

(5-3) Case of  $q = q_1AwBq_2$ , where  $\{A, B\} = \{y_1ay_2, bc\}$  ( $a = d$ ). Suppose that  $(A, B) = (y_1ay_2, bc)$ , i.e.,  $q = q_1y_1ay_2wbcq_2$  holds. Then, the following conditions must be satisfied: For  $y'_1, y'_2 \in X$ ,

- |  |  |
|--|--|
| (1) $p_1 \preceq q_1$ or $p_1 \preceq q_1y'_1$ | (1') $p_2 \preceq y_2wbcq_2$ ,                             |
| (2) $p_1 \preceq q_1y_1ay_2$ ,                 | (2') $p_2 \preceq wbcq_2$ or<br>$p_2 \preceq y'_2wbcq_2$ , |
| (3) $p_1 \preceq q_1y_1ay_2w$ ,                | (3') $p_2 \preceq q_2$ .                                   |

Let  $q'_1 = q_1y_1a$ ,  $q'_2 = y_2w$ ,  $q'_3 = bcq_2$ . From (3) and (1'), we have  $p_1 \preceq q'_1q'_2$  and  $p_2 \preceq q'_2q'_3$ , respectively. Since  $q'_2$  contains a variable symbol, Theorem 2 implies that  $p \preceq q$  holds. This contradicts the assumption. Similarly, we can show that the case  $(A, B) = (bc, y_1ay_2)$  also contradicts the assumption.

From the above, we conclude that if  $p\{x := r\} \preceq q$  for all  $r = \{ya, bc, dy\}$  ( $b \notin \{a, d\}$  and  $c \notin \{a, d\}$ ), then  $p\{x := xy\} \preceq q$  holds.  $\square$

The condition in Lemma 5 is illustrated in four cases (3)-(6) in Fig. 11.

**Lemma 6:** Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$  and let  $p, q$

$bc \in D \wedge p\{a := by\} \not\preceq q$   
 $b \neq d \preceq q$  tr.  $b = c \in D \Rightarrow a = c$   $p\{a := yc\} = p\{x := ya\} \leq q$   
maximally generalized  $b \neq c$  5.2

be regular patterns on  $\Sigma \cup X$ . Let  $D$  be one of the following sets of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol in  $X$  that does not appear in  $p$  and  $q$ .

$$D = \{ya, bc, dy\} \quad (b = a, b \neq d, \text{ and } c \notin \{a, d\})$$

Then, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$ .

We note that if  $b = d$ , then, because  $p\{x := dy\} \preceq q$ ,  $p\{x := bc\} \preceq q$  is always satisfied. In this sense,  $D$  essentially consists of only two elements. To avoid this, we assume  $b \neq d$  D is maximally generalized (F) (N) S)

**Proof.** We assume that  $p\{x := xy\} \not\preceq q$  in order to derive a contradiction. The proof is almost the same as the proof of Lemma 5. Since  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , there are three strings of length 2 corresponding to  $ya, bc, dy$  in  $q$ . The symbols appearing in  $D$  correspond to either a variable or a constant symbol in  $q$ . Let  $y_1$  and  $y_2$  be variable symbols appearing in  $q$ . The strings  $ya$  and  $dy$  must correspond to the strings  $y_1a$  and  $dy_2$  in  $q$ , respectively. For the same reasons stated at the beginning of Lemma 5, the string  $bc$  corresponds to the string  $bc$  in  $q$  as well. Let  $A, B, C$  be regular patterns on  $\Sigma \cup X$ , where  $\{A, B, C\} = \{y_1a, ac, dy_2\}$ . Since  $p\{x := xy\} \not\preceq q$ ,  $q$  can be expressed in one of the following four forms: Let  $y_1, y_2$  be distinct variable symbols in  $X$ , and  $q_1, q_2, w, w'$  either the empty string or a regular pattern on  $\Sigma \cup X$ . From the conditions  $b = a$  and  $b \neq d$ , it follows that  $a \neq d$ .

- (6-1)  $q = q_1AwBw'Cq_2$ ,  
where  $\{A, B, C\} = \{y_1a, ac, dy_2\}$ ,
- (6-2)  $q = q_1AwBq_2$ ,  
where  $\{A, B\} = \{y_1ac, dy_2\}$ ,
- (6-3)  $q = q_1Aq_2$ , where  $A = dy_1ac$ .

In cases (6-1) and (6-2), similar to Lemma 5, it is shown that  $q = q_1y_1awacw'dy_2q_2$  and  $q = q_1y_1acwdy_2q_2$ , respectively, where  $w$  and  $w'$  contain no variable symbols.

(6-1) Case of  $q = q_1AwBw'Cq_2$ , where  $\{A, B, C\} = \{y_1a, ac, dy_2\}$ : The following conditions must be satisfied:

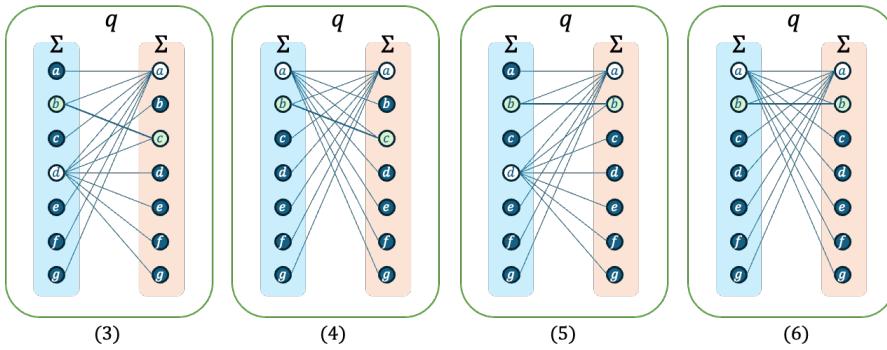
- |                                  |                                   |
|----------------------------------|-----------------------------------|
| (1) $p_1 \preceq q_1$ ,          | (1') $p_2 \preceq wacw'dy_2q_2$ , |
| (2) $p_1 \preceq q_1y_1aw$ ,     | (2') $p_2 \preceq w'dy_2q_2$ ,    |
| (3) $p_1 \preceq q_1y_1awacw'$ , | (3') $p_2 \preceq q_2$ .          |

From (1') and (2'),  $wacw'd$  and  $w'd$  are prefixes of  $p_2$ , and from (2) and (3),  $awacw'$  and  $aw$  are suffixes of  $p_1$ . It implies a contradiction in the following inductive way:

- $|w| = |w'|: c = a$  holds.
- $|w| = |w'| + 1: w = w'd = cw'$  holds. Thus, from Proposition 1,  $c = d$  holds.
- $|w| = |w'| + 2: w = w'da = acw'$  holds. From Proposition 2,  $c \in \{a, d\}$  holds.
- $|w| \geq |w'| + 3:$ From (2) and (3), there exists a string  $w''$  of length  $|w| - |w'| - 2$  such that  $w = w''acw'$  holds. Moreover, from (2) and (3), since  $|aw| < |wacw'|$  and  $aw = aw''acw'$ , it follows that  $aw''$  is a suffix of  $w$ . On the other hand, from (1') and (2'),  $w'd$  is a prefix of  $w$ .

$D = \{ya, ay, bb\}$

$\pi_1 = aab\pi$   
 $\pi_2 = aayb\pi$  L( $\pi_2$ )CL( $\pi_1$ )



**Fig. 11** Let  $\Sigma = \{a, b, c, d, e, f, g\}$  and  $p, q \in \mathcal{RP}$ . We assume that the symbols in  $\Sigma$  are mutually distinct. The figure (3) expresses case  $D = \{ya, bc, dy\}$  in Lemma 5. The figures (4), (5), and (6) express three cases  $D = \{ya, bc, ay\}$ ,  $D = \{ya, bb, dy\}$ , and  $D = \{ya, bb, ay\}$ , respectively. In these cases, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$  holds.

ya

$p\{x := ay\} =$	$e   a   b   c   b   c   a   d   a   b   c   b   c   a   d   a   y   a   b   c   a   d   a   d   a   b   c   b   c   a   d   a   d   e$
	$y_1 \quad \boxed{a} \quad y_2$
$p\{x := bc\} =$	$e   a   b   c   b   c   a   d   a   b   c   b   c   a   d   a   \boxed{b}   c   b   c   a   d   a   d   a   b   c   b   c   a   d   a   d   e$
	$y_1 \quad \boxed{b} \quad \boxed{c} \quad y_2$
$p\{x := dy\} =$	$e   a   b   c   b   c   a   d   a   b   c   b   c   a   d   a   \boxed{d}   y   b   c   a   d   a   d   a   b   c   b   c   a   d   a   d   e$
	$y_1 \quad \boxed{d} \quad y_2$

**Fig. 12** Substitutions for  $p$  and each correspondence to  $q$ .

P.9で参照  
順序が  
おかしいかも

Since  $|w'd| + |aw''| = |w'| + |w''| + 2 = |w|$ , we have  $w = w'daw''$ . Therefore,  $w'daw'' = w''acw'$  holds (Fig. 13). From Proposition 3, we have  $c \in \{a, d\}$ .

- $|w'| = |w| + 1$ : From (1') and (2'),  $c = d$  holds.
- $|w'| = |w| + 2$ : From (1') and (2'),  $d$  is a prefix of  $w'$ . Thus, from (2) and (3),  $w' = wac = daw$  holds. From Proposition 2,  $c \in \{a, d\}$  holds.
- $|w'| \geq |w| + 3$ : From (1') and (2'), there exists a string  $w''$  of length  $|w| - |w'| - 2$  such that  $w' = wacw''$  holds. Moreover, from (1') and (2'), since  $|w'd| < |wacw'|$  and  $w'd = wacw''d$ ,  $w'd$  is a prefix of  $w'$ . On the other hand, from (1') and (2'),  $aw'$  is a suffix of  $w'$ . Since  $|w''d| + |aw'| = |w'| + |w| + 2 = |w'|$ , we have  $w' = w''daw$ . Therefore,  $w''daw = wacw''$  holds.
- From Proposition 3, we have  $c \in \{a, d\}$ .

All the cases contradict the condition  $c \notin \{a, d\}$ . Therefore, if  $b = a$ ,  $b \neq d$ , and  $c \notin \{a, d\}$  are satisfied, case (6-1) is impossible.

(6-2) Case of  $q = q_1AwBq_2$ , where  $\{A, B\} = \{y_1ac, dy_2\}$ : For  $q = q_1y_1acwady_2q_2$ , the following conditions must be satisfied:

- (1)  $p_1 \preceq q_1$ ,
- (2)  $p_1 \preceq q_1y_1$ ,
- (3)  $p_1 \preceq q_1y_1acwad$ ,
- If  $|w| = 0$ , from (1') and (2'), the prefix of  $p_2$  is  $cd$  and

REVISIT!

- (1')  $p_2 \preceq cwdy_3q_2$ ,
- (2')  $p_2 \preceq wdyy_3q_2$ ,
- (3')  $p_2 \preceq q_2$ .

$a   \quad w$	(2)
$a   \quad w \quad   a   c   \quad w'$	(3)
$a   w''   a   c   \quad w'$	
$w \quad   a   c   \quad w'$	(1')
$w' \quad   d$	(2')

$|w| \geq |w'|$

$a   \quad w$	(2)
$a   \quad w \quad   a   c   \quad w'$	(3)
$w \quad   a   c   \quad w''   d$	
$w \quad   a   c   \quad w'$	(1')
$w' \quad   d$	(2')

$|w| < |w'|$

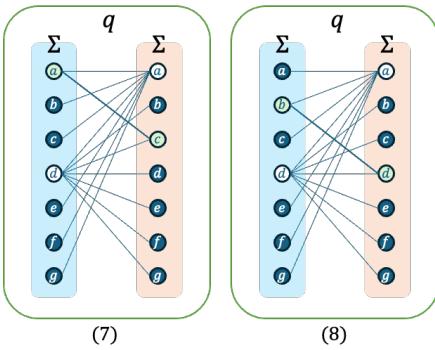
**Fig. 13** Case (6-1) in Lemma 6: Relation of strings  $w$ ,  $w'$ , and  $w''$

他の場合も証明を書くべきかと

- $d$ . Thus, we have  $c = d$ .
- If  $|w| = 1$ , from (1') and (2'), the prefix of  $p_2$  is  $cwd$  and  $wd$ . Thus, we have  $w = c = d$ .
- If  $|w| \geq 2$ , then from (1') and (2'),  $cwd$  and  $wd$  are prefixes of  $p_2$ . Thus, we have  $cw = wd$ . From Proposition 2,  $c = d$  holds.

All of these cases do not meet  $b = a$ ,  $b \neq d$ , and  $c \notin \{a, d\}$ . Therefore, if  $b = a$ ,  $b \neq d$ , and  $c \notin \{a, d\}$  are satisfied, case

集合でよろしく?



**Fig. 14** Let  $\Sigma = \{a, b, c, d, e, f, g\}$  and  $p, q \in \mathcal{RP}$ . We assume that the symbols in  $\Sigma$  are mutually distinct. The figures (7) and (8) express two cases  $D = \{ya, ac, dy\}$  and  $D = \{ya, bd, dy\}$  in Lemmas 6 and 7, respectively. In these cases, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$  holds.

(6-2) is also impossible.

(6-3) Case of  $q = q_1 A q_2$ , where  $A = dy_1 ac$ : For  $q = q_1 dy_1 ac q_2$ , the following conditions must be satisfied for  $y'_1, y''_1 \in X$ :

- |   |   |
|---|---|
| $(1) p_1 \preceq q_1$<br>$(2) p_1 \preceq q_1 dy_1$ ,<br>$(3) p_1 \preceq q_1,$ | $(1') p_2 \preceq q_2$ ,<br>$(2') p_2 \preceq q_2$ ,<br>$(3') p_2 \preceq acq_2$ or<br>$p_2 \preceq y''_1 acq_2.$ |
|---|---|

For  $p_1 \preceq q_1 d$  in (1) and  $p_2 \preceq acq_2$  in (3'),  $p = p_1 x p_2 \preceq q_1 dx ac q_2 \preceq q\{y_1 := x\}$  holds. From this, we have  $p\{x := xy\} \preceq q\{y_1 := x\}$ . This contradicts the assumption that  $p\{x := xy\} \not\preceq q$ . Similarly, we can show that the other cases of (1) and (3') also contradict the assumption.

From the above, we conclude that if  $p\{x := r\} \preceq q$  for all  $r \in \{ya, bc, dy\}$  ( $b = a$ ,  $b \neq d$ , and  $c \notin \{a, d\}$ ), then  $p\{x := xy\} \preceq q$  holds.  $\square$

The conditions in Lemmas 6 and 7 are illustrated in (7) and (8) in Fig. 14, respectively.

**Lemma 7:** Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$  and let  $p, q$  be regular patterns on  $\Sigma \cup X$ . Let  $D$  be one of the following sets of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol in  $X$  that does not appear in  $p$  and  $q$ .

$$D = \{ya, bc, dy\} \quad (b \notin \{a, d\}, \quad c \neq a \text{ and } c = d).$$

Then, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$ .

**Proof.** The proof follows by reversing  $p$  and  $q$  and subsequently applying Lemma 6.  $\square$

When the conditions of Lemmas 5, 6, and 7 are not satisfied, counterexamples can be constructed as follows:

**Proposition 4:** Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$ . For a variable symbol  $y$ , let  $D = \{ya, bc, dy\}$  ( $b = a$  and  $c = d$ ). There exist regular patterns  $p$  and  $q$  on  $\Sigma \cup X$  such that  $p\{x := r\} \preceq q$  for any  $r \in D$ , but  $p\{x := xy\} \not\preceq q$ .

$$c=d, \quad a \neq c, \quad b \neq d, \quad b \neq c$$

$$a \neq d, \quad \Sigma \ni a, b, c, d \quad (a, b, c, d \in \Sigma)$$

**Proof.** We provide an example to demonstrate this proposition. Let  $a, b, c, d, e$  be constant symbols in  $\Sigma$ , and let  $x, y, y_1, y_2$  be variable symbols in  $X$ . Define the regular patterns  $p$  and  $q$  as follows:

$$\begin{aligned} p &= eabc \underset{\text{red}}{bc} a \underset{\text{red}}{d} abc \underset{\text{red}}{bc} c a d a \underset{\text{red}}{b} c a d a \underset{\text{red}}{d} a b c \underset{\text{red}}{c} b c a d a d e, \\ q &= y_1 a b c \underset{\text{green}}{b c} a \underset{\text{green}}{d} a b c \underset{\text{green}}{b c} c a d a y_2 \quad (b = a \text{ and } c = d). \end{aligned}$$

Obviously  $p\{x := xy\} \not\preceq q$  holds. For these  $p$  and  $q$ , the condition for Proposition 4 holds as follows (see also Fig. 12):

$$\begin{aligned} p \{x := \underset{\text{red}}{y_1}\} &= (eabc \underset{\text{red}}{bc} a \underset{\text{red}}{d} abc \underset{\text{red}}{bc} c a d a \underset{\text{red}}{b} c a d a \underset{\text{red}}{d} a b c \underset{\text{red}}{c} b c a d a d e) \underset{\text{green}}{q} \\ &= q\{y_1 := eabc \underset{\text{green}}{bc} a \underset{\text{green}}{d} abc \underset{\text{green}}{bc} c a d a, y_2 := e\} \\ &\preceq q, \\ p \{x := b c\} &= (eabc \underset{\text{green}}{c} a \underset{\text{green}}{d} abc \underset{\text{green}}{b} c a d a \underset{\text{green}}{d} a b c \underset{\text{green}}{c} b c a d a d e) \underset{\text{green}}{q} \\ &= q\{y_1 := eabc \underset{\text{green}}{c} a \underset{\text{green}}{d} abc \underset{\text{green}}{b} c a d a, y_2 := a b c \underset{\text{green}}{c} a d a d e\} \\ &\preceq q, \\ p \{x := d y\} &= eabc \underset{\text{green}}{c} a \underset{\text{green}}{d} abc \underset{\text{green}}{b} c a d a \underset{\text{green}}{d} (y b c a d a \underset{\text{green}}{a} b c \underset{\text{green}}{c} b c a d a d e) \underset{\text{green}}{q} \\ &= q\{y_1 := e, y_2 := y b c a d a \underset{\text{green}}{a} b c \underset{\text{green}}{c} b c a d a d e\} \\ &\preceq q. \end{aligned}$$

□

### 3.4 $D = \{a_1 b_1, a_2 b_2, a_3 y\}$ and $D = \{a_1 b_1, a_2 b_2, y b_3\}$

**Lemma 8:** Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$  and  $p, q$  regular patterns on  $\Sigma \cup X$ . Let  $D$  be the following set of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol in  $X$  that does not appear in  $p$  and  $q$ :

$$D = \{a_1 b_1, a_2 b_2, a_3 y\},$$

where  $a_i \neq a_j$  and  $b_i \neq b_j$  ( $i \neq j, 1 \leq i, j \leq 3$ ).

可階性  $\Sigma \geq 3$   $\Sigma \geq 3$   
 $a_1 \neq a_2, a_2 \neq a_3, a_1 \neq a_3$   
 $b_1 \neq b_2, b_2 \neq b_3, b_1 \neq b_3$

Then, if  $p\{x := r\} \preceq q$  holds for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$  holds.

**Proof.** We assume that  $p\{x := xy\} \not\preceq q$  holds. Since  $D$  is maximally generalized on  $(p, q)$ , from the same argument as in the proof of Lemma 6, it is sufficient to consider the following five cases (8-1)–(8-5) of  $q$ : For  $y_1 \in X$ ,

- (8-1)  $q = q_1 a_1 b_1 w a_2 b_2 w' a_3 y_1 q_2$ ,
- (8-2)  $q = q_1 a_1 b_1 b_2 y_1 q_2$  ( $a_2 = b_1$  and  $a_3 = b_2$ ),
- (8-3)  $q = q_1 a_1 b_1 b_2 w a_3 y_1 q_2$  ( $b_1 = a_2$ ),
- (8-4)  $q = q_1 a_3 y_1 w a_1 b_1 b_2 q_2$  ( $b_1 = a_2$ ),
- (8-5)  $q = q_1 a_1 b_1 y_1 w a_2 b_2 q_2$  ( $b_1 = a_3$ ),

where no variable symbol appears in both  $w$  and  $w'$ .

(8-1) Case of  $q = q_1 a_1 b_1 w a_2 b_2 w' a_3 y_1 q_2$ : The following conditions must be satisfied: For  $y'_1 \in X$ ,

- (1)  $p_1 \preceq q_1$ , (1')  $p_2 \preceq w a_2 b_2 w' a_3 y_1 q_2$ ,
- (2)  $p_1 \preceq q_1 a_1 b_1 w$ , (2')  $p_2 \preceq w' a_3 y_1 q_2$ ,
- (3)  $p_1 \preceq q_1 a_1 b_1 w a_2 b_2 w'$ , (3')  $p_2 \preceq q_2$  or  $p_2 \preceq y'_1 q_2$ .

$|w|=|w'|$  の場合がでてどうなるかはま。

いきなり説明入るってほんとくねえ

△説明入るってほんとくねえ

$$a_2 b_2 w' = a_2 b_2 w_1 w$$

the last symbol of  $w$  is  $a_3$ . Thus,  $a_1 = a_3$  holds. This contradicts the assumption of  $a_1 \neq a_3$ .

(8-4) Case of  $q = q_1 a_3 y_1 w a_1 b_1 b_2 q_2$  ( $b_1 = a_2$ ): The following conditions must be satisfied: For  $y'_1 \in X$ ,

- |                                       |   |
|---------------------------------------|---|
| (1) $p_1 \preceq q_1$ ,               | (1') $p_2 \preceq w a_1 b_1 b_2 q_2$ or<br>$p_2 \preceq y'_1 w a_1 b_1 b_2 q_2$ , |
| (2) $p_1 \preceq q_1 a_3 y_1 w$ ,     | (2') $p_2 \preceq b_2 q_2$ ,  |
| (3) $p_1 \preceq q_1 a_3 y_1 w a_1$ , | (3') $p_2 \preceq q_2$ .  |

From (3), there exist regular patterns  $p'_1$  and  $p''_1$  such that  $p_1 = p'_1 p''_1$ ,  $p'_1 \preceq q_1 a_3$ , and  $p''_1 \preceq y_1 w a_1$  hold. Hence, if  $p_2 \preceq w a_1 b_1 b_2 q_2$  of (1') holds, since  $p = p_1 x p_2 = p'_1 p''_1 x p_2 \preceq q_1 a_3 p''_1 x w a_1 b_1 b_2 q_2 = q\{y_1 := p''_1 x\}$ , then  $p \preceq q$  holds. Thus, this contradicts the assumption. Similarly,  $p_2 \preceq y'_1 w a_1 b_1 b_2 q_2$  of (1') leads to a contradiction.

(8-5) Case of  $q = q_1 a_1 b_1 y_1 w a_2 b_2 q_2$  ( $b_1 = a_3$ ): The following conditions must be satisfied: For  $y'_1 \in X$ ,

- |                                       |   |
|---------------------------------------|---|
| (1) $p_1 \preceq q_1$ ,               | (1') $p_2 \preceq y_1 w a_2 b_2 q_2$ ,                                    |
| (2) $p_1 \preceq q_1 a_1$ ,           | (2') $p_2 \preceq w a_2 b_2 q_2$ or<br>$p_2 \preceq y'_1 w a_2 b_2 q_2$ , |
| (3) $p_1 \preceq q_1 a_1 b_1 y_1 w$ , | (3') $p_2 \preceq q_2$ .  |

Let  $q'_1 = q_1 a_1 b_1$ ,  $q'_2 = y_1 w$ ,  $q'_3 = a_2 b_2 q_2$ . From (3),  $p_1 \preceq q'_1 q'_2$  holds, and from (1'),  $p_2 \preceq q'_2 q'_3$  holds. Since  $q'_2$  contains a variable symbol  $y_1$ ,  $p \preceq q$  holds from Theorem 2. This contradicts the assumption.  $\square$

(8-2) Case of  $q = q_1 a_1 b_1 b_2 y_1 q_2$  ( $a_2 = b_1$  and  $a_3 = b_2$ ): The following conditions must be satisfied: For  $y'_1 \in X$ ,

- |                                 |  |
|---------------------------------|--|
| (1) $p_1 \preceq q_1$ ,         | (1') $p_2 \preceq b_2 y_1 q_2$ ,                   |
| (2) $p_1 \preceq q_1 a_1$ ,     | (2') $p_2 \preceq y_1 q_2$ ,                       |
| (3) $p_1 \preceq q_1 a_1 b_1$ , | (3') $p_2 \preceq q_2$ or $p_2 \preceq y'_1 q_2$ . |

From (2) and (3),  $a_1 b_1$  and  $a_1$  are suffixes of  $p_1$ . Hence,  $b_1 = a_1$  holds. Thus, from the assumption of  $b_1 = a_2$ ,  $a_1 = a_2$  holds. This contradicts the assumption of  $a_1 \neq a_2$ .

(8-3) Case of  $q = q_1 a_1 b_1 b_2 w a_3 y_1 q_2$  ( $b_1 = a_2$ ): The following conditions must be satisfied: For  $y'_1 \in X$ ,

- |                                       |  |
|---------------------------------------|--|
| (1) $p_1 \preceq q_1$ ,               | (1') $p_2 \preceq b_2 w a_3 y_1 q_2$ ,             |
| (2) $p_1 \preceq q_1 a_1$ ,           | (2') $p_2 \preceq w a_3 y_1 q_2$ ,                 |
| (3) $p_1 \preceq q_1 a_1 b_1 b_2 w$ , | (3') $p_2 \preceq q_2$ or $p_2 \preceq y'_1 q_2$ . |

- $|w| = 0$ : From (2) and (3),  $a_1$  and  $a_1 b_1 b_2$  are suffixes of  $p_1$ . Hence,  $a_1 = b_2$  holds. Moreover, since  $b_2 a_3$  and  $a_3$  are prefixes of  $p_2$ ,  $b_2 = a_3$  holds. Thus,  $a_1 = a_3$  holds. This contradicts the assumption of  $a_1 \neq a_3$ .
- $|w| \geq 1$ : Since  $a_1$  and  $a_1 b_1 b_2 w$  are suffixes of  $p_1$  from (2) and (3), the last symbol of  $w$  is  $a_1$ . Moreover, since  $b_2 w a_3$  and  $w a_3$  are prefixes of  $p_2$  from (1') and (2'),

$$D = \{a_1 b_1, a_2 b_2, y b_3\},$$

where  $a_i \neq a_j$  and  $b_i \neq b_j$  ( $i \neq j$ ,  $1 \leq i, j \leq 3$ ).

Then, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$ .

$p\{x := a_1 y\} \neq q$   
 $p\{x := y b_1\} \neq q$   
 $p\{x := a_2 z\} \neq q$   
 $p\{x := y b_2\} \neq q$

**Proof.** The proof follows by reversing  $p$  and  $q$  and subsequently applying Lemma 8.  $\square$

### 3.5 $D = \{a_1 b_1, a_2 b_2, a_3 b_3\}$

In Lemma 14 (ii) of [4], they stated that, when  $\#\Sigma \geq 3$ , for regular patterns  $p, q$ , if  $p\{x := r\} \preceq q$  for any  $r \in D$ , then  $p\{x := xy\} \preceq q$  holds, where  $D = \{a_1 b_1, a_2 b_2, a_3 b_3\}$  ( $a_i \neq a_j$  and  $b_i \neq b_j$  for each  $i, j$  ( $i \neq j$ ,  $1 \leq i, j \leq 3$ )). Unfortunately, there exist the following counterexamples of Lemma 14 (ii) of [4].

CR 動子法で証明

**Example 2:** Assume that  $a_1 = b_2$  and  $a_3 = b_1$  hold.

- (1) Let  $p = c a_1 x' a_3 c$  and  $q = x a_1 a_3 y$ . It is clear that  $\{x := xy\} \not\preceq q$  holds. However, we can see that  $p\{x' := a_1 b_1\} \preceq q$ ,  $p\{x' := a_2 b_2\} \preceq q$  and  $p\{x' := a_3 b_3\} \preceq q$

P71 なぜ?

$$a_1=b_2, a_2=b_1 \quad q = a_1 a_1 b_3 y \\ = a_1 b_1 y$$

$$ca_1 y b_2 b_1 c = ca_1 y a_1 b_1 c$$

定数  
P定数  
する元  
hold, since  $p\{x' := a_1 b_1\} = ca_1 a_1 b_1 a_3 c = q\{x := ca_1, y := a_3 c\}$ ,  $p\{x' := a_2 b_2\} = ca_1 a_2 b_2 a_3 c = q\{x := ca_1 a_2, y := c\}$  and  $p\{x' := a_3 b_3\} = ca_1 a_3 b_3 a_3 c = q\{x := c, y := b_3 a_3 c\}$  hold.

- (2) Let  $p = cb_2 a_1 b_1 b_2 x' a_1 b_1 b_2 a_3 c$  and  $q = xb_2 a_1 b_1 b_2 a_3 y$ . It is clear that  $p\{x = xy\} \not\leq q$  holds. However, we have  $p\{x' := a_1 b_1\} \leq q$ ,  $p\{x' := a_2 b_2\} \leq q$ , and  $p\{x' := a_3 b_3\} \leq q$ , since  $p\{x' := a_1 b_1\} = cb_2 a_1 b_1 b_2 a_1 b_1 b_2 a_3 c = q\{x := cb_2 a_1 b_1, y := b_2 a_3 c\}$ ,  $p\{x' := a_2 b_2\} = cb_2 a_1 b_1 b_2 a_2 b_2 a_1 b_1 b_2 a_3 c = q\{x := cb_2 a_1 b_1 b_2 a_2, y := c\}$ , and  $p\{x' := a_3 b_3\} = cb_2 a_1 b_1 b_2 a_3 b_3 a_1 b_1 b_2 a_3 c = q\{x := c, y := b_3 a_1 b_1 b_2 a_3 c\}$  hold.

The conditions in Lemmas 8, 9, and 10 are illustrated in the cases (9), (10), and (11) in Fig. 15.

**Lemma 10:** Let  $\Sigma$  be an alphabet with  $\#\Sigma \geq 3$  and  $p, q$  regular patterns on  $\Sigma \cup X$ . Let  $D$  be the following set of regular patterns on  $\Sigma \cup X$ , where  $y$  is a variable symbol in  $X$  that does not appear in  $p$  and  $q$ :

$$D = \{a_1 b_1, a_2 b_2, a_3 b_3\},$$

where  $a_i \neq a_j$  and  $b_i \neq b_j$  ( $i \neq j, 1 \leq i, j \leq 3$ )

Then, if  $p\{x := r\} \leq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \leq q$ .

**Proof.** We assume that  $p\{x := xy\} \not\leq q$  holds. Since  $D$  is maximally generalized on  $(p, q)$ , it is sufficient to consider the following four cases (10-1)-(10-4) of  $q$  for some regular patterns  $q_1, q_2$  and some constant strings  $w, w'$  ( $|w| \geq 0$  and  $|w'| \geq 0$ ):

- (10-1)  $q = q_1 a_1 b_1 w a_2 b_2 w' a_3 b_3 q_2$ ,
- (10-2)  $q = q_1 a_1 b_1 a_3 b_3 q_2$  ( $b_1 = a_2$  and  $a_3 = b_2$ ),
- (10-3)  $q = q_1 a_1 b_1 b_2 w a_3 b_3 q_2$  ( $b_1 = a_2$ ),
- (10-4)  $q = q_1 a_1 b_1 w a_2 b_2 b_3 q_2$  ( $b_2 = a_3$ ).

(10-1) Case of  $q = q_1 a_1 b_1 w a_2 b_2 w' a_3 b_3 q_2$ : The following conditions must be satisfied:

- (1)  $p_1 \leq q_1$ , (1')  $p_2 \leq w a_2 b_2 w' a_3 b_3 q_2$ ,
- (2)  $p_1 \leq q_1 a_1 b_1 w$ , (2')  $p_2 \leq w' a_3 b_3 q_2$ ,
- (3)  $p_1 \leq q_1 a_1 b_1 w a_2 b_2 w'$ , (3')  $p_2 \leq q_2$ .

- $|w| = |w'|$ : From (2) and (3),  $a_1 b_1 w a_2 b_2 w'$  and  $a_1 b_1 w$  are suffixes of  $p_1$ . Then,  $a_1 b_1 w = a_2 b_2 w'$ . Hence,  $a_1 b_1 = a_2 b_2$ . This contradicts the assumption of  $a_1 \neq a_2$  and  $b_1 \neq b_2$ . ( $|w| + 1 = |w'|$ ) The strings  $w a_2 b_2 w' a_3 b_3$  and  $w' a_3 b_3$  are prefixes of  $p_2$ . If there exists a constant symbol  $w_1$  such that  $w' a_3 b_3 = w w_1 a_3 b_3$ , then  $b_2$  and  $a_3$  are the same symbol from  $w a_2 b_2 = w w_1 a_3$ . From (2) and (3),  $a_1 b_1 w a_2 b_2 w'$  and  $a_1 b_1 w$  are suffixes of  $p_1$ . Then, there exists a constant symbol  $w_2$  such that  $w' = w_2 w$ , then  $b_2$  and  $a_1$  are the same symbol from  $w_2 w a_2 = a_1 b_1 w$ . Hence, from  $b_2 = a_3$ ,  $a_3$  and  $a_1$  are same symbol. This contradicts the assumption of  $a_3 \neq a_1$ .
- $|w| + 1 < |w'|$ : From (2) and (3),  $a_1 b_1 w a_2 b_2 w'$  and

$a_1 b_1 w$  are suffixes of  $p_1$ . If there exists a constant string  $w_1$  ( $|w_1| \geq 2$ ) such that  $w' = w_1 w$ , then  $a_1 b_1$  is a suffix of  $w_1$ . From conditions (1') and (2'),  $w a_2 b_2 w' a_3 b_3$  and  $w' a_3 b_3$  are prefixes of  $p_2$ . If there exist constant strings  $w_1$  and  $w_2$  such that  $w' = w_1 w = w w_2$  holds, then  $w w_2 a_3 b_3$  are suffixes of  $w_1$  from  $|w_1| = |w_2|$  and  $w w_2 a_3 b_3 = w a_2 b_2 w_1$ . Hence,  $a_1 b_1 = a_3 b_3$ . This contradicts the assumption of  $a_1 \neq a_3$  and  $b_1 \neq b_3$ .

- $|w| > |w'|$ : We can prove the contradiction in a similar way as  $|w| \leq |w'|$ .

(10-2) Case of  $q = q_1 a_1 b_1 a_3 b_3 q_2$  ( $b_1 = a_2$  and  $a_3 = b_2$ ): The following conditions must be satisfied:

- |                              |                               |
|------------------------------|-------------------------------|
| (1) $p_1 \leq q_1$ ,         | (1') $p_2 \leq a_3 b_3 q_2$ , |
| (2) $p_1 \leq q_1 a_1$ ,     | (2') $p_2 \leq b_3 q_2$ ,     |
| (3) $p_1 \leq q_1 a_1 b_1$ , | (3') $p_2 \leq q_2$ .         |

From (2) and (3), since  $a_1 b_1$  and  $a_1$  are suffixes of  $p_1$ ,  $b_1 = a_1$  holds. From the assumption of  $b_1 = a_2$ , the equation  $a_1 = a_2$  holds. This contradicts the assumption of  $a_1 \neq a_2$ .

(10-3) Case of  $q = q_1 a_1 b_1 b_2 w a_3 b_3 q_2$  ( $b_1 = a_2$ ): The following conditions must be satisfied:

- |                                    |                                     |
|------------------------------------|-------------------------------------|
| (1) $p_1 \leq q_1$ ,               | (1') $p_2 \leq b_2 w a_3 b_3 q_2$ , |
| (2) $p_1 \leq q_1 a_1$ ,           | (2') $p_2 \leq w a_3 b_3 q_2$ ,     |
| (3) $p_1 \leq q_1 a_1 b_1 b_2 w$ , | (3') $p_2 \leq q_2$ .               |

- $|w| = 0$ : From (2) and (3),  $a_1$  and  $a_1 b_1 b_2$  are suffixes of  $p_1$ . Moreover, from (1') and (2'),  $b_2 a_3 b_3$  and  $a_3 b_3$  are prefixes of  $p_2$ . Since  $b_2 = a_1$  and  $b_2 a_3 = a_3 b_3$ ,  $a_1 = a_3$  holds. This contradicts the assumption of  $a_1 \neq a_3$ .

- $|w| \geq 1$ : From (2) and (3),  $a_1$  and  $a_1 b_1 b_2 w$  are suffixes of  $p_1$ . Hence, the last symbol of  $w$  is  $a_1$ . Moreover,  $b_2 w a_3 b_3$  and  $w a_3 b_3$  are prefixes of  $p_2$  from (1') and (2'). Hence, the last symbol of  $w$  is  $a_3$ . Therefore,  $a_1 = a_3$  holds. This contradicts the assumption of  $a_1 \neq a_3$ .

(10-4) Case of  $q = q_1 a_1 b_1 w a_2 b_2 b_3 q_2$  ( $b_2 = a_3$ ): The following conditions must be satisfied:

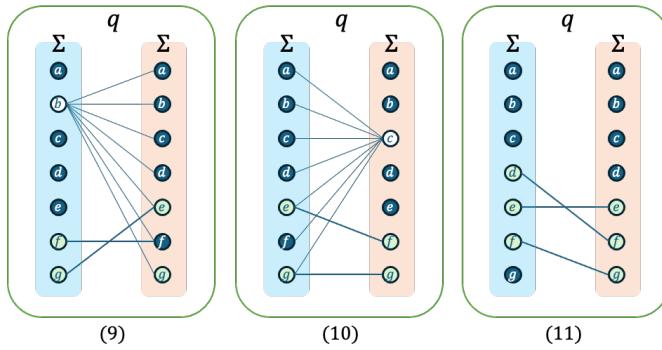
- |                                    |                                     |
|------------------------------------|-------------------------------------|
| (1) $p_1 \leq q_1$ ,               | (1') $p_2 \leq w a_2 b_2 b_3 q_2$ , |
| (2) $p_1 \leq q_1 a_1 b_1 w$ ,     | (2') $p_2 \leq b_3 q_2$ ,           |
| (3) $p_1 \leq q_1 a_1 b_1 w a_2$ , | (3') $p_2 \leq q_2$ .               |

- $|w| = 0$ : From (2) and (3),  $a_1 b_1$  and  $a_1 b_1 a_2$  are suffixes of  $p_1$ . And from (1') and (2'),  $a_2 b_2 b_3$  and  $b_3$  are prefixes of  $p_2$ . Since  $b_1 = a_2$  and  $a_2 = b_3$ , then  $b_1 = b_3$  holds. This contradicts the assumption of  $b_1 \neq b_3$ .

- $|w| \geq 1$ : Since  $a_1 b_1 w$  and  $a_1 b_1 w a_2$  are suffixes of  $p_1$  from (2) and (3), the first symbol of  $w$  is  $b_1$ . Moreover, since  $w a_2 b_2 b_3$  and  $b_3$  are prefixes of  $p_2$  from (1') and (2'), the first symbol of  $w$  is  $b_3$ . Therefore,  $b_1 = b_3$  holds. This contradicts the assumption of  $b_1 \neq b_3$ .

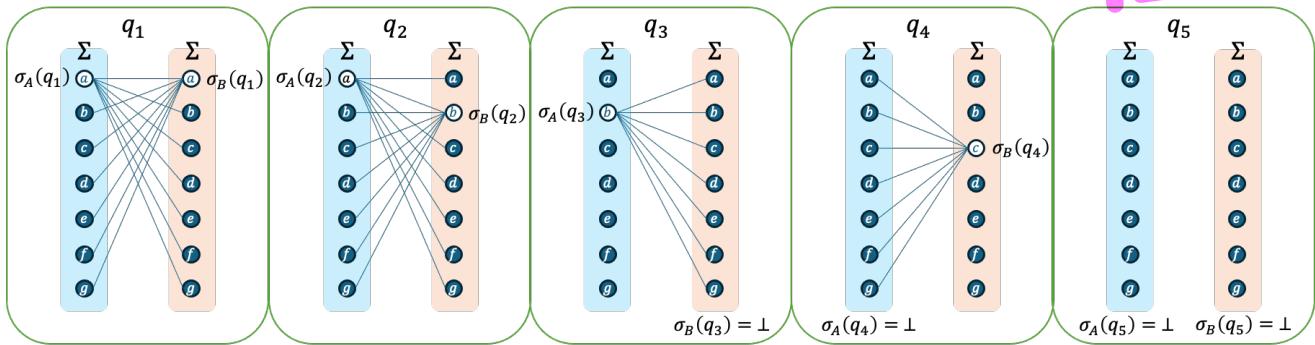
□

$a_1 = a_3$  などとやつて  
ページの省略にてます



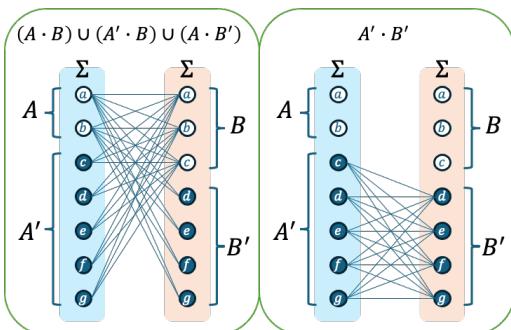
**Fig. 15** Let  $\Sigma = \{a, b, c, d, e, f, g\}$  and  $p, q \in \mathcal{RP}$ . We assume that the symbols in  $\Sigma$  are mutually distinct. The figures (9), (10), and (11) express cases  $D = \{a_1b_1, a_2b_2, a_3y\}$ ,  $D = \{a_1b_1, a_2b_2, yb_3\}$ , and  $D = \{a_1b_1, a_2b_2, a_3b_3\}$  in Lemmas 8, 9, and 10, respectively, where  $a_i \neq a_j$  and  $b_i \neq b_j$  for each  $i, j$  ( $i \neq j, 1 \leq i, j \leq 3$ ). In these cases, if  $p\{x := r\} \preceq q$  for all  $r \in D$  and  $D$  is maximally generalized on  $(p, q)$ , then  $p\{x := xy\} \preceq q$  holds.

$a_1=f, b_1=f_1$   
 $a_2=g, b_2=e$   
 $a_3=c \rightarrow f \rightarrow b?$   
 $\Sigma = \{a, b, \dots, g\} \subseteq \Sigma$   
 開始はやがくがよく  
 つかうね...



**Fig. 16** Let  $\Sigma = \{a, b, c, d, e, f, g\}$ ,  $Q = \{q_1, q_2, q_3, q_4, q_5\}$ . We set  $A(q_1) = \{a\}$  and  $B(q_1) = \{a\}$ , and then  $\sigma_A(q_1) = a$  and  $\sigma_B(q_1) = a$ , and so on. For each regular pattern  $q_i$  ( $i = 1, \dots, 5$ ), we represent a string  $w \in \Sigma \cdot \Sigma$  satisfying that  $p\{x := w\} \preceq q_i$  by the edge between the left (first) and right (second) symbols of  $w$ . For example, the leftmost figure shows that  $p\{x := ay\} \preceq q_1$  and  $p\{x := ya\} \preceq q_1$  for a variable symbol  $y$ . We note that these figures may contain more edges than those illustrated. From these figures, we get  $\ell_A = 1$ ,  $\ell_B = 0$ , and  $Q^{(\perp, \perp)} = \{q_5\}$ ,  $Q^{(\perp, \cdot)} = \{q_4\}$ ,  $Q^{(\cdot, \perp)} = \{q_3\}$ ,  $Q^{(\cdot, \cdot)} = \{q_1, q_2\}$ .

S2の条件が満たす。



**Fig. 17** In the left figure, we aggregate all of the edges appearing in Fig. 16. For all  $w = a'b' \in A' \cdot B'$ , there must be a regular pattern  $q_i$  ( $1 \leq i \leq 5$ ) that satisfies that  $p\{x := w\} \preceq q_i$ .

alphabet with  $\#\Sigma = k + 2$ . Let  $p \in \mathcal{RP}$  in which a variable symbol  $x$  appears, and let  $Q \in \mathcal{RP}^k$ . If for any string  $w \in \Sigma^*$  with  $|w| = 2$ , there exists a regular pattern  $q_w \in Q$  such that  $p\{x := w\} \preceq q_w$  holds, then there exists a regular pattern  $q \in Q$  such that  $p\{x := xy\} \preceq q$  holds, where  $y$  is a variable symbol that does not appear in  $q$ .

**Proof.** Without loss of generality, we suppose that  $\#Q = k$  holds. Otherwise, for some regular pattern  $q$  already in  $Q$ , we can add a new regular pattern  $q'$  equivalent to  $q$ , i.e.,  $q' \equiv q$ , to  $Q$  repeatedly until  $\#Q = k$  is satisfied. For any  $q \in Q$ , we define the sets  $A(q), B(q) \subseteq \Sigma$  as follows:

$$\begin{aligned} A(q) &= \{a \in \Sigma \mid p\{x := ay\} \preceq q, y \in X\}, \\ B(q) &= \{b \in \Sigma \mid p\{x := yb\} \preceq q, y \in X\}. \end{aligned}$$

If there exists  $q \in Q$  such that  $|A(q)| \geq 2$  or  $|B(q)| \geq 2$ , from Lemma 4,  $p\{x := xy\} \preceq q$  holds. Below, we suppose that  $|A(q)| \leq 1$  and  $|B(q)| \leq 1$ . Let  $\perp$  be a constant symbol that is not a member in  $\Sigma$ . We define the functions  $\sigma_A : Q \rightarrow \Sigma \cup \{\perp\}$  and  $\sigma_B : Q \rightarrow \Sigma \cup \{\perp\}$  as follows:

### 3.6 Characteristic sets for finite union of regular patterns

**Lemma 11:** Let  $k$  be an integer with  $k \geq 1$ . Let  $\Sigma$  be an

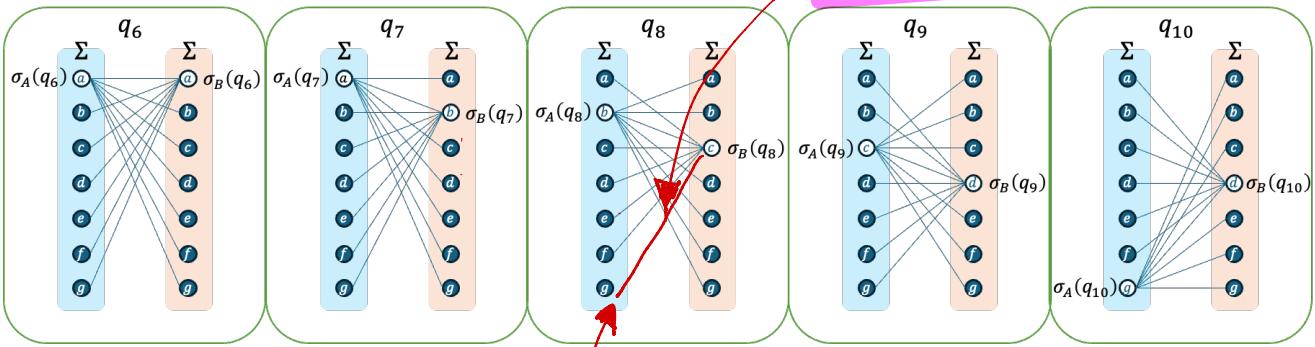


Fig. 18 Let  $\Sigma = \{a, b, c, d, e, f, g\}$ ,  $Q = \{q_6, q_7, q_8, q_9, q_{10}\}$ . From these figures, we get  $\ell_A = 1$ ,  $\ell_B = 1$ ,  $Q^{(\perp, \perp)} = Q^{(\perp, \cdot)} = Q^{(\cdot, \perp)} = \emptyset$ , and  $Q^{(\cdot, \cdot)} = Q$ .

$\alpha_A^{-1}(a) \subset Q$   
 $\alpha_B^{-1}(a) \subset Q$

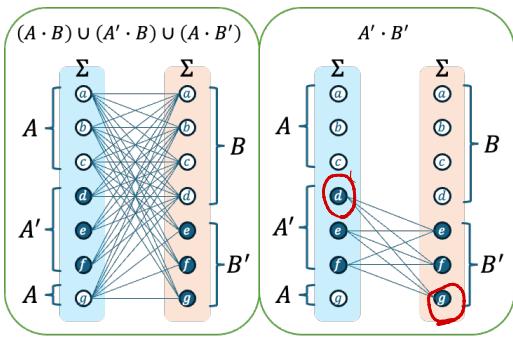


Fig. 19 In the left figure, we aggregate all of the edges appearing in Fig. 18. From Fig. 18 and this right figure, we get  $Q_1^{(\cdot, \cdot)} = \{q_6, q_7, q_8, q_9\}$  and  $Q_2^{(\cdot, \cdot)} = \{q_{10}\}$ . From Proposition 4, even if the string  $dg \in A' \cdot B'$  satisfies  $p\{x := gd\} \leq q_{10}$ , it does not imply that  $p\{x := xy\} \leq q_{10}$ .

$$\sigma_A(q) = \begin{cases} a & \text{if } A(q) = \{a\}, \\ \perp & \text{if } A(q) = \emptyset. \end{cases}$$

$$\sigma_B(q) = \begin{cases} b & \text{if } B(q) = \{b\}, \\ \perp & \text{if } B(q) = \emptyset. \end{cases}$$

The inverse functions of  $\sigma_A$  and  $\sigma_B$  are denoted by  $\sigma_A^{-1}$  and  $\sigma_B^{-1}$ , respectively. That is, for  $a, b \in \Sigma \cup \{\perp\}$ , let  $\sigma_A^{-1}(a) = \{q \in Q \mid \sigma_A(q) = a\}$  and  $\sigma_B^{-1}(b) = \{q \in Q \mid \sigma_B(q) = b\}$ . We give an example in Fig. 16.

A and B denotes the following subsets of  $\Sigma$ :

$$A = \bigcup_{q \in Q \setminus \sigma_A^{-1}(\perp)} A(q), \quad B = \bigcup_{q \in Q \setminus \sigma_B^{-1}(\perp)} B(q).$$

Then, let  $A' = \Sigma \setminus A$  and  $B' = \Sigma \setminus B$ . For any  $a, b \in \Sigma$ , we use the following notations:

$$\ell_A = \sum_{a \in A} (\#\sigma_A^{-1}(a) - 1), \quad \ell_B = \sum_{b \in B} (\#\sigma_B^{-1}(b) - 1).$$

These  $\ell_A$  and  $\ell_B$  represent the numbers of excess duplicate symbols in A and B. We easily see the following claim:

Claim 1.

$$(i) \#A + \#A' = \#B + \#B' = k + 2,$$

$$(ii) \#A + \ell_A + \#\sigma_A^{-1}(\perp) = \#B + \ell_B + \#\sigma_B^{-1}(\perp) = k.$$

Since  $\#\Sigma = k + 2$  and  $\#Q = k$ ,  $\#A' \geq 2$  and  $\#B' \geq 2$  hold. We partition Q into the following subsets:

$$Q^{(\perp, \perp)} = \sigma_A^{-1}(\perp) \cap \sigma_B^{-1}(\perp),$$

$$Q^{(\perp, \cdot)} = \sigma_A^{-1}(\perp) \cap (Q \setminus \sigma_B^{-1}(\perp)),$$

$$Q^{(\cdot, \perp)} = (Q \setminus \sigma_A^{-1}(\perp)) \cap \sigma_B^{-1}(\perp),$$

$$Q^{(\cdot, \cdot)} = (Q \setminus \sigma_A^{-1}(\perp)) \cap (Q \setminus \sigma_B^{-1}(\perp)).$$

Lemma 2.8

From the condition of this lemma, for any string  $w \in \Sigma^*$  with  $|w| = 2$ , there exists a regular pattern  $q_w \in Q$  such that  $p\{x := w\} \leq q_w$  holds. In particular, for  $w = a'b' \in A' \cdot B'$ , we must have  $q_w \in Q$  that satisfies that  $p\{x := w\} \leq q_w$  (Fig. 17). It is easy to see that if  $w \in (A \cdot B) \cup (A' \cdot B) \cup (A \cdot B')$ , there exists a regular pattern  $q_w \in Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)} \cup Q^{(\cdot, \cdot)}$  such that  $p\{x := w\} \leq q_w$  holds. We have the following two claims:

A' · B' is maximally generalized prefix

Claim 2. If there exist  $q \in Q^{(\perp, \perp)}$  and distinct 5 strings  $w_i \in A' \cdot B'$  ( $1 \leq i \leq 5$ ) such that  $p\{x := w_i\} \leq q$  holds ( $1 \leq i \leq 5$ ), then  $p\{x := xy\} \leq q$  holds.

Proof of Claim 2. Let  $W = \{a_1b_1, \dots, a_5b_5\} \subset A' \cdot B'$ . Because, for any  $i$  ( $1 \leq i \leq 5$ ),  $|W \cap \{a_i c \mid c \in \Sigma\}| \leq 2$  and  $|W \cap \{c b_i \mid c \in \Sigma\}| \leq 2$ , it can be proven that there are 3 strings  $a_{i_1}b_{i_1}, a_{i_2}b_{i_2}, a_{i_3}b_{i_3} \in W$  such that  $a_{i_j} \neq a_{i_{j'}}$  and  $b_{i_j} \neq b_{i_{j'}}$  for any  $i_j, i_{j'} (i_j \neq i_{j'}, 1 \leq j, j' \leq 3)$ . Therefore, from Lemma 10, this claim holds. (End of Proof of Claim 2)

Claim 3. If there exist  $q \in Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)}$  and distinct 3 strings  $w_i \in A' \cdot B'$  ( $1 \leq i \leq 3$ ) such that  $p\{x := w_i\} \leq q$  holds ( $1 \leq i \leq 3$ ), then  $p\{x := xy\} \leq q$  holds.

Proof of Claim 3. Let  $W = \{a_1b_1, a_2b_2, a_3b_3\} \subset A' \cdot B'$ . Because, for any  $i$  ( $1 \leq i \leq 3$ ),  $|W \cap \{a_i c \mid c \in \Sigma\}| \leq 2$  and  $|W \cap \{c b_i \mid c \in \Sigma\}| \leq 2$ , it can be proven that there are 2 strings  $a_{i_1}b_{i_1}, a_{i_2}b_{i_2} \in W$  such that  $a_{i_1} \neq a_{i_2}$  and  $b_{i_1} \neq b_{i_2}$ . Therefore, from Lemmas 8 and 9, this claim holds. (End of Proof of Claim 3)

If there exist a regular pattern  $q \in Q^{(\perp, \perp)} \cup Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)}$  and enough strings  $w \in A' \cdot B'$  such that either of the conditions of Claims 2 and 3 is satisfied, this lemma holds. Then, we assume that it is not the case.

Therefore

**Assumption 1.** There is no regular pattern  $q \in Q^{(\perp, \perp)}$  and 5 strings  $w \in A' \cdot B'$  such that the condition of *Claim 2* is satisfied and there is no regular pattern  $q \in Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)}$  and 3 strings  $w \in A' \cdot B'$  such that the condition of *Claim 3* is satisfied.

Let  $\mathcal{L}_1 = \#\{w \in A' \cdot B' \mid \exists q \in Q^{(\perp, \perp)} \cup Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)} \text{ s.t. } p\{x := w\} \preceq q\}$ . Under *Assumption 1*, each  $q \in Q^{(\perp, \perp)}$  has at most 4 strings  $w \in A' \cdot B'$  such that the condition of *Claim 2* is satisfied, and each  $q \in Q^{(\perp, \cdot)} \cup Q^{(\cdot, \perp)}$  has at most 2 strings  $w \in A' \cdot B'$  such that the condition of *Claim 3* is satisfied. Then, by *Claim 1*,

$$\begin{aligned}\mathcal{L}_1 &\leq 4\#Q^{(\perp, \perp)} + 2\#Q^{(\perp, \cdot)} + 2\#Q^{(\cdot, \perp)} \\ &= 2(\#Q^{(\perp, \perp)} + \#Q^{(\perp, \cdot)}) + 2(\#Q^{(\perp, \perp)} + \#Q^{(\cdot, \perp)}) \\ &= 2\#\sigma_A^{-1}(\perp) + 2\#\sigma_B^{-1}(\perp) \\ &= 2(k - \#A - \ell_A) + 2(k - \#B - \ell_B) \\ &= 2(\#A' - \ell_A - 2) + 2(\#B' - \ell_B - 2) \\ &= 2(\#A' + \#B') - 2(\ell_A + \ell_B) - 8.\end{aligned}$$

Next, we partition  $Q^{(\cdot, \cdot)}$  into the following two subsets:

$$\begin{aligned}Q_1^{(\cdot, \cdot)} &= \{q \in Q^{(\cdot, \cdot)} \mid \sigma_A(q) \in B \text{ or } \sigma_B(q) \in A\}, \\ Q_2^{(\cdot, \cdot)} &= \{q \in Q^{(\cdot, \cdot)} \mid \sigma_A(q) \in B' \text{ and } \sigma_B(q) \in A'\}.\end{aligned}$$

We show the next two claims on  $Q_1^{(\cdot, \cdot)}$  and  $Q_2^{(\cdot, \cdot)}$ .

**Claim 4.** If there exist  $q \in Q_1^{(\cdot, \cdot)}$  and a string  $a'b' \in A' \cdot B'$  such that  $p\{x := a'b'\} \preceq q$  holds, then  $p\{x := xy\} \preceq q$  holds.

*Proof of Claim 4.* Suppose that both  $\sigma_A(q) \in B$  and  $\sigma_B(q) \in A$  hold. Then, since  $a' \notin \{\sigma_A(q), \sigma_B(q)\} \subseteq A \cap B$  and  $b' \notin \{\sigma_A(q), \sigma_B(q)\} \subseteq A \cap B$ , from Lemma 5,  $p\{x := xy\} \preceq q$  holds. Suppose that  $\sigma_A(q) \in B$  and  $\sigma_B(q) \in A'$ . If  $a' = \sigma_B(q)$ , since  $a' \in B$ ,  $a' \neq b'$  holds. Since  $\sigma_A(q) \in B$ ,  $b' \neq \sigma_A(q)$  holds. That is,  $a' = \sigma_B(q)$ ,  $a' \neq \sigma_A(q)$ , and  $b' \notin \{\sigma_A(q), \sigma_B(q)\}$  hold. Therefore, from Lemmas 6 and 7,  $p\{x := xy\} \preceq q$  holds. If  $a' \neq \sigma_B(q)$ , since  $b' \neq \sigma_A(q)$ , from Lemma 5,  $p\{x := xy\} \preceq q$  holds. Similarly, the case that  $\sigma_A(q) \in B'$  and  $\sigma_B(q) \in A$  is proven. (End of Proof of Claim 4)

**Claim 5.** If there exist  $q \in Q_2^{(\cdot, \cdot)}$  and a string  $a'b' \in A' \cdot B'$  such that  $(a' \neq \sigma_B(q) \text{ or } b' \neq \sigma_A(q))$  and  $p\{x := a'b'\} \preceq q$  hold, then  $p\{x := xy\} \preceq q$  holds.

*Proof of Claim 5.* When  $a' = b'$ , since  $\sigma_A(q) \neq \sigma_B(q)$ , from Lemma 5, this claim holds. Similarly, when  $a' \neq b'$ , from Lemmas 5, 6, and 7, this holds. (End of Proof of Claim 5)

We give an example in Fig. 18 and Fig. 19.

If there exist a regular pattern  $q \in Q_2^{(\cdot, \cdot)}$  and a string  $w \in A' \cdot B'$  such that the condition of *Claim 5* is satisfied, this lemma holds. Then, we also assume that it is not the case.

**Assumption 2.** There is no  $q \in Q_2^{(\cdot, \cdot)}$  and a string  $a'b' \in A' \cdot B'$  such that the condition of *Claim 5* is satisfied.

Let  $\mathcal{L}_2 = \#\{a'b' \in A' \cdot B' \mid \exists q \in Q_2^{(\cdot, \cdot)} \text{ s.t. } p\{x := a'b'\} \preceq q\}$ . For any  $a'b' \in A' \cdot B'$  and  $q \in Q_2^{(\cdot, \cdot)}$ , if  $a' = \sigma_B(q)$  and

$$\begin{array}{llll}a' \notin B & a' \in A' & a' \in A & a' \in B \\ b' \in B & b' \notin B & b' \in B & b' \notin B\end{array}$$

Since  $a' \notin \{\sigma_A(q), \sigma_B(q)\}$

and  $b' \notin \{\sigma_A(q), \sigma_B(q)\}$  for  $i = 1, 2$

$$\begin{array}{ll}A' = \sigma_B(q) & p\{x := xy\} \preceq q \\ B' = \sigma_A(q) & p\{x := xy\} \preceq q\end{array}$$

$b' = \sigma_A(q)$  hold (it is the condition of Proposition 4), by considering the duplicate numbers  $\ell_A$  and  $\ell_B$ , we have the following inequality:

$$\mathcal{L}_2 \leq \min\{\#A' + \ell_B, \#B' + \ell_A\}.$$

We show the last claim:

$$\text{Claim 6. } \#A' \times \#B' - \mathcal{L}_1 - \mathcal{L}_2 \geq 2.$$

*Proof of Claim 6.* First we prove the inequality when  $\#A \leq k-1$  and  $\#B \leq k-1$ , i.e.,  $\#A' \geq 3$  and  $\#B' \geq 3$  hold. Since  $\mathcal{L}_2 \leq \frac{1}{2}(\#A' + \#B' + \ell_A + \ell_B)$ ,

$$\begin{aligned}&\#A' \times \#B' - \mathcal{L}_1 - \mathcal{L}_2 \\ &\geq \#A' \times \#B' - (2(\#A' + \#B') - 2(\ell_A + \ell_B) - 8) \\ &\quad - \frac{1}{2}(\#A' + \#B' + \ell_A + \ell_B) \\ &= \#A' \times \#B' - \frac{5}{2}(\#A' + \#B') + \frac{3}{2}(\ell_A + \ell_B) + 8 \\ &= (\#A' - \frac{5}{2})(\#B' - \frac{5}{2}) + \frac{3}{2}(\ell_A + \ell_B) + \frac{7}{4} \geq 2.\end{aligned}$$

$$\frac{25}{4} + \frac{25}{4} = \frac{25}{2}$$

When  $\#A = k$  and  $\#B \leq k$ , i.e.,  $\#A' = 2$  and  $\#B' \geq 2$  hold, since  $\ell_A = 0$ ,  $\mathcal{L}_1 \leq 2\#B' - 2\ell_B - 4$  holds. Moreover,  $\mathcal{L}_2 \leq \min\{\#B', \ell_B + 2\}$  holds. From *Claim 1*,  $\ell_B + 2 = k - \#\sigma_B^{-1}(\perp) - \#B = \#B' - \#\sigma_B^{-1}(\perp)$  holds. Therefore,  $\mathcal{L}_2 \leq \ell_B + 2$  holds. Thus,

$$\begin{aligned}&\#A' \times \#B' - \mathcal{L}_1 - \mathcal{L}_2 \\ &\geq 2\#B' - (2\#B' - 2\ell_B - 4) - (\ell_B + 2) \\ &= \ell_B + 2 \geq 2.\end{aligned}$$

Similarly, the case when  $\#A \leq k$  and  $\#B = k$  is proven. (End of Proof of Claim 6)

Under Assumptions 1 and 2, from *Claim 6*, there exist at least two  $w \in A' \cdot B'$  and a regular pattern  $q \in Q_1^{(\cdot, \cdot)}$  such that the condition of *Claim 4* is satisfied. Therefore, for such a regular pattern  $q$ ,  $p\{x := xy\} \preceq q$  holds.  $\square$

**Lemma 12** (Sato et al.[4]): Let  $\Sigma$  be a finite alphabet with  $\#\Sigma \geq 3$  and  $p, q$  regular patterns. If there exists a constant symbol  $a \in \Sigma$  such that  $p\{x := a\} \preceq q$  and  $p\{x := xy\} \preceq q$ , then  $p \preceq q$  holds, where  $y$  is a variable symbol that does not appear in  $q$ .

From Lemma 11 and Lemma 12, we have the following theorem.

**Theorem 4:** Let  $k \geq 3$ ,  $\#\Sigma \geq 2k-1$ ,  $P \in \mathcal{RP}^+$  and  $Q \in \mathcal{RP}^k$ . Then, the following (i), (ii) and (iii) are equivalent:

- (i)  $S_2(P) \subseteq L(Q)$ , (ii)  $P \sqsubseteq Q$ , (iii)  $L(P) \subseteq L(Q)$ .

**Proof.** It is clear that (ii) implies (iii) and (iii) implies (i). From Theorem 3, if  $\#\Sigma \geq 2k+1$ , then (i) implies (ii). Let  $\#Q = k$ ,  $p \in P$ ,  $\#\Sigma = 2k-1$  or  $2k$ . Then, we show that (i) implies (ii). It suffices to show that  $S_2(p) \subseteq L(Q)$  implies  $P \sqsubseteq Q$  for any regular pattern  $p \in \mathcal{RP}$ . The proof is done by mathematical induction on  $n$ , where  $n$  is the number of

$P \in \mathcal{P}$

variable symbols appears in  $p$ .

In case  $n = 0$ ,  $S_2(p) = \{p\}$  holds. By (i), we have  $\{p\} \subseteq L(Q)$ . Thus,  $p \preceq q$  for some  $q \in Q$ .

For  $n \geq 0$ , we assume that it is valid for any regular pattern  $p$  with  $n$  variable symbols. Let  $p$  be a regular pattern such that  $n + 1$  variable symbols appear in  $p$  and  $S_2(p) \subseteq L(Q)$ . We assume that  $p \not\preceq q_i$ , that is,  $p \not\preceq q_i$  for any  $i \in \{1, \dots, k\}$ . Let  $Q = \{q_1, \dots, q_k\}$  and  $p_1, p_2$  regular patterns,  $x$  a variable symbol with  $p = p_1xp_2$ . For  $a, b \in \Sigma$ , let  $p_a = p\{x := a\}$  and  $p_{ab} = p\{x := ab\}$ . Both  $p_a$  and  $p_{ab}$  have  $n$  variable symbols, respectively. Thus,  $S_2(p_a) \subseteq L(Q)$  and  $S_2(p_{ab}) \subseteq L(Q)$  hold. By the induction hypothesis, there exist  $i, i' \in \{1, \dots, k\}$  such that  $p_a \preceq q_i$  and  $p_{ab} \preceq q_{i'}$ . Let  $D_i = \{a \in \Sigma \mid p\{x := a\} \preceq q_i\}$  ( $i = 1, \dots, k$ ). We assume that  $\#D_i \geq 3$  for some  $i \in \{1, \dots, k\}$ . By Lemma 2, we have  $p \preceq q_i$ . This contradicts the assumption. Thus, we have  $\#D_i \leq 2$  for any  $i \in \{1, \dots, k\}$ . If  $\#\Sigma = 2k - 1$ , then  $\#D_i = 2$  or  $\#D_i = 1$  for any  $i \in \{1, \dots, k\}$ . Moreover, If  $\#\Sigma = 2k$ , then  $\#D_i = 2$  for any  $i \in \{1, \dots, k\}$ . Since  $k \geq 3$ ,  $2k + 1 \geq k + 2$  holds. By Lemma 11, there exists  $i \in \{1, \dots, k\}$  such that  $p\{x := xy\} \preceq q_i$ . Therefore, by Lemma 12, we have  $p \preceq q_i$ . This contradicts the assumption. Thus, (i) implies (ii).  $\square$

From Theorem 4, the following Corollary 2 holds.

**Corollary 2:** Let  $k \geq 3$ ,  $\#\Sigma \geq 2k - 1$  and  $P \in \mathcal{RP}^+$ . Then,  $S_2(P)$  is a characteristic set for  $L(P)$  within  $\mathcal{RPL}^k$ .

**Lemma 13** (Sato et al.[4]): Let  $k \geq 3$  and  $\#\Sigma \leq 2k - 2$ . Then,  $\mathcal{RP}^k$  does not have compactness with respect to containment.

$$P = a_1a_1w_1w_1b_1b_1, \quad q_1 = a_1a_1w_1b_1b_1$$

**Proof.** Let  $\Sigma = \{a_1, \dots, a_{k-1}, b_1, \dots, b_{k-1}\}$  and  $p, q_i$  regular patterns,  $w_i \in \Sigma^*$  ( $i = 1, \dots, k - 1$ ) defined in a similar way to Example 1. Let  $q_k = x_1a_1w_1xyw_1b_1x_2$ . Since  $p\{x := a_i\} = x_1a_1w_1a_iw_1b_1x_2 \preceq q_i$  and  $p\{x := b_i\} = x_1a_1w_1b_iw_1b_1x_2 \preceq q_i$  for any  $i \in \{1, \dots, k - 1\}$ , we have  $S_1(p) \subseteq \bigcup_{i=1}^{k-1} L(q_i)$ . For any  $w \in \{s \in \Sigma^+ \mid |s| \geq 2\}$ ,  $p\{x := w\} = x_1a_1w_1ww_1b_1x_2 \preceq q_k$ . Thus, we have  $L(p) \subseteq L(Q)$ . By Theorem 1, since  $p \not\preceq q_i$ ,  $L(p) \not\subseteq L(q_i)$  for any  $i \in \{1, \dots, k\}$ . Therefore,  $\mathcal{RP}^k$  does not have compactness with respect to containment.  $\square$

From Theorem 4 and Lemma 13, we have the following Theorem 5.

**Theorem 5:** Let  $k \geq 3$  and  $\#\Sigma \geq 2k - 1$ . Then,  $\mathcal{RP}^k$  has compactness with respect to containment.

In case  $k = 2$ , we have the following theorem.

**Theorem 6:** Let  $\#\Sigma \geq 4$ ,  $P \in \mathcal{RP}^+$  and  $Q \in \mathcal{RP}^2$ . The following (i), (ii) and (iii) are equivalent:

- (i)  $S_2(P) \subseteq L(Q)$ ,
- (ii)  $P \sqsubseteq Q$ ,
- (iii)  $L(P) \subseteq L(Q)$ .

**Proof.** It is clear that (ii) implies (iii), and (iii) implies (i). Thus, we show that (i) implies (ii). It suffices to show that  $S_2(p) \subseteq L(Q)$  implies  $P \sqsubseteq Q$  for any regular pattern  $p \in \mathcal{RP}$ . Let  $Q = \{q_1, q_2\}$ . The proof is done by mathematical induction on  $n$ , where  $n$  is the number of variable symbols

appearing in  $p$ . In case  $n = 0$ ,  $p \in \Sigma^+$ . Since  $S_2(p) = \{p\} \subseteq L(Q)$ , we have  $p \preceq q$  for some  $q \in Q$ . For  $n \geq 0$ , we assume that it is valid for any regular pattern  $p$  with  $n$  variable symbols. Let  $p$  be a regular pattern such that  $n + 1$  variable symbols appear in  $p$ , and  $S_2(p) \subseteq L(Q)$  holds. We assume that  $p \not\preceq q_i$  ( $i = 1, 2$ ). Let  $p_1, p_2$  be regular patterns and  $x$  a variable symbol with  $p = p_1xp_2$ . For  $a, b \in \Sigma$ , let  $p_a = p\{x := a\}$  and  $p_{ab} = p\{x := ab\}$ . Note that  $p_a$  and  $p_{ab}$  have  $n$  variable symbols. Thus, by the assumption,  $S_2(p_a) \subseteq L(Q)$  and  $S_2(p_{ab}) \subseteq L(Q)$  imply  $p_a \preceq q_i$  and  $p_{ab} \preceq q_{i'}$  for some  $i, i' \in \{1, 2\}$ . Let  $D_i = \{a \in \Sigma \mid p\{x := a\} \preceq q_i\}$  ( $i = 1, 2$ ). By Lemma 2, if  $\#D_i \geq 3$  for some  $i \in \{1, 2\}$ , then  $p \preceq q_i$ . This contradicts that  $p \not\preceq q_i$  ( $i = 1, 2$ ). Thus, we have  $\#D_i \leq 2$  for any  $i \in \{1, 2\}$ . Since  $\#\Sigma \geq 4$ , we consider that  $\#D_1 = 2$  and  $\#D_2 = 2$ . From Lemma 11,  $p\{x := xy\} \preceq q_i$  for some  $i \in \{1, 2\}$ . From Lemma 12, we have  $p \preceq q_i$  for some  $i \in \{1, 2\}$ . This contradicts that  $p \not\preceq q_i$  ( $i = 1, 2$ ). Hence, (i) implies (ii).  $\square$

The next example gives a set of regular patterns  $P \in \mathcal{RP}^+$  and a set of regular patterns  $Q \in \mathcal{RP}^2$  that, in case  $\#\Sigma = 3$ , the three conditions (i), (ii) and (iii) in Theorem 6 are not equivalent?

**Example 3:** Let  $\Sigma = \{a, b, c\}$ ,  $p, q_1, q_2$  regular patterns and  $x, x', x''$  variable symbols such that  $p = x'axbx''$ ,  $q_1 = x'abx''$  and  $q_2 = x'cx''$ . Let  $w \in \Sigma^+$ . If  $w$  contains  $c$ , then  $p\{x := w\} \preceq q_2$ . On the other hand, if  $w$  does not contain  $c$ , then  $p\{x := w\} \preceq q_1$ . Thus,  $L(p) \subseteq L(q_1) \cup L(q_2)$ . However,  $p \not\preceq q_1$  and  $p \not\preceq q_2$ .

From Theorem 6, the following two corollaries holds.

**Corollary 3:** Let  $\#\Sigma \geq 4$  and  $P \in \mathcal{RP}^+$ . Then,  $S_2(P)$  is a characteristic set for  $L(P)$  within  $\mathcal{RPL}^2$ .

**Corollary 4:** Let  $\#\Sigma \geq 4$ . Then,  $\mathcal{RP}^2$  has compactness with respect to containment.

#### 4. Regular Pattern without Adjacent Variable Symbols

A regular pattern  $p$  is said to be a *non-adjacent variable regular pattern* (NAV regular pattern) if  $p$  does not contain consecutive variable symbols. For example, the regular pattern  $p = axybc$  is not an NAV regular pattern because  $xy$  is appeared in  $p$ . Let  $\mathcal{RP}_{\text{NAV}}$  be the set of all NAV regular patterns. Let  $\mathcal{RP}_{\text{NAV}}^+$  be the set of all finite subsets  $S$  of  $\mathcal{RP}_{\text{NAV}}$  such that  $S$  is not the empty set, i.e.,  $\mathcal{RP}_{\text{NAV}}^+ = \{S \subseteq \mathcal{RP}_{\text{NAV}} \mid \#S \leq 1\}$ , and  $\mathcal{RP}_{\text{NAV}}^k$  the set of all subsets  $P$  of  $\mathcal{RP}_{\text{NAV}}^+$  such that  $P$  consists of at most  $k$  ( $k \geq 1$ ) NAV regular patterns, i.e.,  $\mathcal{RP}_{\text{NAV}}^k = \{P \in \mathcal{RP}_{\text{NAV}}^+ \mid \#P \leq k\}$ . We define the compactness with respect to containment for  $\mathcal{RP}_{\text{NAV}}^k$  in a similar way as Def 2. For any NAV regular pattern  $p \in \mathcal{RP}_{\text{NAV}}$  and any set  $Q \in \mathcal{RP}_{\text{NAV}}^k$  with  $k$  ( $k \geq 1$ ), the set  $\mathcal{RP}_{\text{NAV}}^k$  said to have *compactness with respect to containment* if there exists an NAV regular pattern  $q \in Q$  such that  $L(p) \subseteq L(q)$  holds if  $L(p) \subseteq L(Q)$  holds. Then, the following Theorem 7 holds.

**Theorem 7:** For an integer  $k$  ( $k \geq 2$ ), let  $\#\Sigma \geq k + 2$ ,  $P \in$

この定理は複雑で、手書きの証明が複数箇所に散在する。

$p = x'cadadaadacbadadaadaxadadaaadabadaadabx'',$   
 $q_1 = x'cadadaadacbadadaadacx'',$   
 $q_2 = x'badadaadacx'',$   
 $q_3 = x'aadadx''.$

Fig. 20 NAV regular patterns  $p, q_1, q_2$ , and  $q_3$

$\mathcal{RP}_{NAV}^+, Q \in \mathcal{RP}_{NAV}^k$ . Then, the following (i), (ii) and (iii) are equivalent:

- (i)  $S_2(P) \subseteq L(Q)$ , (ii)  $P \sqsubseteq Q$ , (iii)  $L(P) \subseteq L(Q)$ .

**Proof.** From the definitions of  $\mathcal{RP}_{NAV}^+$  and  $\mathcal{RP}_{NAV}^k$ , it is clear that (ii) implies (iii) and (iii) implies (i). Hence, we will show that (i) implies (ii) by mathematical induction on the number **n** of variable symbols that appear in an NAV regular pattern  $p \in P$  as follows: If  $n = 0$ , then we have  $S_2(\{p\}) = \{p\}$ . Hence,  $p \in L(Q)$ . Therefore, there exists  $q \in Q$  such that  $p \preceq q$  holds.

If  $n \geq 0$ , we assume that the proposition holds for any regular NAV regular pattern containing  $n \geq 0$  variable symbols. Let  $p$  be an NAV regular pattern containing  $n + 1$  variable symbols such that  $S_2(\{p\}) \subseteq L(Q)$  holds and  $p$  contains a variable symbol  $x$ . There exist two NAV regular patterns  $p_1, p_2$  such that  $p = p_1xp_2$  holds. By the induction hypothesis, for any constant string  $w \in \Sigma^*$  with  $|w| = 2$ ,  $\{p\{x := w\}\} \sqsubseteq Q$  holds because  $p\{x := w\}$  contains  $n$  variable symbols. Hence, there exists an NAV regular pattern  $q_w \in Q$  such that  $p\{x := w\} \preceq q_w$  holds. From Lemma 11, there exists a regular pattern  $q \in Q$  such that  $p\{x := xy\} \preceq q$  holds, where  $y$  is a variable symbol that does not appear in  $q$ . This contradicts the condition  $Q \in \mathcal{RP}_{NAV}^k$ . Thus, we have that (i) implies (ii).  $\square$

**Corollary 5:** Let  $k \geq 2$ ,  $\#\Sigma \geq k + 2$  and  $P \in \mathcal{RP}_{NAV}^+$ . Then,  $S_2(P)$  is a characteristic set of  $\mathcal{RP}_{NAV}^k$ .

**Lemma 14:** Let  $k \geq 2$  and  $\#\Sigma \leq k + 1$ . Then,  $\mathcal{RP}_{NAV}^k$  does not have compactness with respect to containment.

**Proof.** Let  $\Sigma = \{a_1, \dots, a_{k+1}\}$ . We assume that for  $i = 1, 2, \dots, k$ ,  $p\{x := a_iy\} \preceq q_i$  and  $p\{x := ya_{i+1}\} \preceq q_i$  holds. If  $p\{x := a_{k+1}a_1\} \preceq q_1$  holds,  $S_2(p) \setminus S_1(p) \subseteq \bigcup_{i=1}^k L(q_i)$  holds. This shows that  $L(p) \subseteq L(Q)$  holds. However, for  $i = 1, 2, \dots, k$ , since  $p \not\preceq q_i$  holds, we have that  $L(p) \not\subseteq L(q_i)$  holds. Hence,  $\mathcal{RP}_{NAV}^k$  does not have compactness with respect to containment.  $\square$

Next, in Example 4, we give an example for Lemma 14.

**Example 4:** Let  $\Sigma$  be the set of four constant symbols  $a, b, c, d$ , i.e.,  $\Sigma = \{a, b, c, d\}$  and  $x, x', x''$  three distinct variable symbols. Let  $p, q_1, q_2, q_3$  be the NAV regular patterns given in Fig. 20. Then, we have  $L(p) \subseteq L(q_1) \cup L(q_2) \cup L(q_3)$ . This shows that for  $P = \{p\}$ ,  $Q = \{q_1, q_2, q_3\}$ , (iii) of Theorem 7 holds. However, since  $p \not\preceq q_1$ ,  $p \not\preceq q_2$  and  $p \not\preceq q_3$  hold, we have  $P \not\sqsubseteq Q$ , that is, (ii) of Theorem 7 does not hold.

From Theorem 7 and Lemma 14, we have the following theorem.

**Table 2** The conditions on the number  $\#\Sigma$  of constant symbols in  $\Sigma$  required for compactness with respect to containment.

Class	$k = 2$	$k \geq 3$
$\mathcal{RP}^k$	$\#\Sigma \geq 4$	$\#\Sigma \geq 2k - 1$
$\mathcal{RP}_{NAV}^k$		$\#\Sigma \geq k + 2$

**Theorem 8:** Let  $k \geq 2$  and  $\#\Sigma \geq k + 2$ . Then, the set  $\mathcal{RP}_{NAV}^k$  has compactness with respect to containment.

## 5. Conclusion

In this paper, for an integer  $k$  ( $k \geq 2$ ), we have shown the conditions on the number of constant symbols in  $\Sigma$ , summarized in Table 2, required for the classes  $\mathcal{RP}^k$  of all the set of  $k$  regular pattern languages and  $\mathcal{RP}_{NAV}^k$  of all the set of  $k$  non-adjacent variable regular patterns in  $\mathcal{RP}_{NAV}$  to have compactness with respect to containment. This result leads to design an efficient learning algorithm for finite unions of languages of non-adjacent variable regular patterns in  $\mathcal{RP}_{NAV}$ , based on the learning algorithm for  $\mathcal{RP}^k$  proposed by Arimura et al. [8].

Extending the notion of strong compactness, as introduced by Arimura et al. [9], to finite unions of regular pattern languages with non-adjacent variables remains as a topic for future research. Furthermore, based on the characteristic set for  $\mathcal{RP}_{NAV}^k$ , we plan to propose a polynomial-time inductive inference algorithm that identifies finite unions of regular pattern languages with non-adjacent variables in the limit from positive examples. Ishinada et al. [17] investigated a query learning model that employs high-precision Graph Convolution Networks (GCNs) as oracles for tree patterns. Applying the findings of the present study to tree pattern languages, with the aim of enabling the extension of their work to finite unions of tree pattern languages, remains an important direction for future research.

## Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Numbers JP20K04973, JP21K12021, JP22K12172, JP24K15074, and JP24K15090. We thank Mr. K. Horii, a master's student at the Graduate School of Information Sciences, Hiroshima City University, for fruitful discussions and constructive comments.

## References

- [1] D. Angluin, "Finding Patterns Common to a Set of Strings," Journal of Computer and System Sciences, 21(1):46–62, 1980, DOI:10.1016/0022-0000(80)90041-0.
- [2] D. Angluin, "Inductive Inference of Formal Languages from Positive Data," Information and Control, 45(2):117–135, 1980, DOI:10.1016/S0019-9958(80)90285-5.
- [3] Y. Mukouchi, "Characterization of Pattern Languages," in Proc. ALT '91, Ohmusha, pp.93-104, 1991.
- [4] D. Sato, Y. Mukouchi and D. Zheng, "Characteristic Sets for Unions of Regular Pattern Languages and Compactness," in Proc. ALT '98,

- Springer LNAI 1501, pp.220-233, 1998.
- [5] Y. Mukouchi, "Containment Problems for Pattern Languages," IEICE Transactions on Information and Systems, E75-D(4):420-425, 1992.
  - [6] K. Wright, "Identification of Unions of Languages Drawn from an Identifiable Class," in Proc. COLT 1989, pp.328-333, 1989.
  - [7] T. Shinohara and H. Arimura, "Inductive inference of unbounded unions of pattern languages from positive data," Theoretical Computer Science, 241(1-2): 135-161, 2000, DOI:10.1016/S0304-3975(99)00270-4.
  - [8] H. Arimura, T. Shinohara and S. Otsuki, "Finding Minimal Generalizations for Unions of Pattern Languages and Its Application to Inductive Inference from Positive Data," in Proc. STACS '94, Springer LNCS 775, pp.649-660, 1994.
  - [9] H. Arimura and T. Shinohara, "Strong Compactness of Containment for Unions of Regular Pattern Languages (in Japanese)," RIMS Kôkyûroku of Kyoto Univ., Vol.950, pp.246-249, 1996.
  - [10] J.D. Day, D. Reidenbach and M.L. Schmid, "Closure Properties of Pattern Languages," Journal of Computer and System Sciences 84:11-31, 2017, DOI:10.1016/j.jcss.2016.07.003.
  - [11] S. Matsumoto, T. Uchida, T. Shoudai, Y. Suzuki, and T. Miyahara, "An Efficient Learning Algorithm for Regular Pattern Languages Using One Positive Example and a Linear Number of Membership Queries," IEICE Trans. Inf. & Syst., vol.E103-D, No.3, pp.526-539, 2020, DOI:10.1587/transinf.2019FCP0009.
  - [12] N. Taketa, T. Uchida, T. Shoudai, S. Matsumoto, Y. Suzuki, and T. Miyahara, "Visualizing the Prediction Basis of Deep Learning Models using a Query Learning Algorithm for Linear Pattern Languages (in Japanese)", JSAI2022(The 36th Annual Conference of the Japanese Society for Artificial Intelligence), 2G4-GS-2-03, 2022.
  - [13] S. Arikawa, T. Shinohara, and A. Yamamoto, "Learning elementary formal system," Theoretical Computer Science, vol.95, no.1, pp.97-113, 1992, DOI:10.1016/0304-3975(92)90068-Q
  - [14] H. Arimura, H. Ishizaka and T. Shinohara, "Learning Unions of Tree Patterns Using Queries," Theor. Comput. Sci., vol.185, No.1, pp.47-62, 1997, DOI:10.1016/S0304-3975(97)00015-7.
  - [15] Y. Suzuki, T. Shoudai, T. Uchida, and T. Miyahara, "Ordered Term Tree Languages Which are Polynomial Time Inductively Inferable from Positive Data," Theoretical Computer Science, vol.350, No.1, pp.63-90, 2006, DOI:10.1016/j.tcs.2005.10.022.
  - [16] T. Uchida, T. Shoudai, and S. Miyano, "Parallel Algorithms for Refutation Tree Problem on Formal Graph Systems," IEICE Trans. Inf. & Syst., vol.E78-D, No.2 pp.99-112, 1995.
  - [17] K. Ishinada, T. Shoudai, T. Uchida, and S. Matsumoto, "Analysis of Query Learning Models with High-Accuracy GCN Oracles for Unordered Tree Patterns (in Japanese)," IPSJ SIG Technical Report on Mathematical Modeling and Problem Solving (MPS), 15, 2023.

**Tomoyuki Uchida** received the B.S. degree in Mathematics, the M.S. and Dr. Sci. degrees in Information Systems all from Kyushu University, in 1989, 1991 and 1994, respectively. Currently, he is a professor of Graduate School of Information Sciences, Hiroshima City University. His research interests include data mining from semistructured data, algorithmic graph theory and algorithmic learning theory.

**Takayoshi Shoudai** received the B.S. in 1986, the M.S. degree in 1988 in Mathematics and the Dr. Sci. in 1993 in Information Science all from Kyushu University. Currently, he is a professor of Department of Computer Science and Engineering, Fukuoka Institute of Technology. His research interests include graph algorithms, computational learning theory and machine learning.

**Satoshi Matsumoto** is a professor of Department of Mathematical Sciences, Tokai University, Kanagawa, Japan. He received the B.S. degree in Mathematics, the M.S. and Dr. Sci. degrees in Information Systems all from Kyushu University, Fukuoka, Japan in 1993, 1995 and 1998, respectively. His research interests include algorithmic learning theory.

**Yusuke Suzuki** received the B.S. degree in Physics, the M.S. and Dr. Sci. degrees in Informatics all from Kyushu University, in 2000, 2002 and 2007, respectively. He is currently a lecturer of Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan. His research interests include machine learning and data mining.

**Tetsuhiro Miyahara** is an associate professor of Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan. He received the B.S. degree in Mathematics, the M.S. and Dr. Sci. degrees in Information Systems all from Kyushu University, Fukuoka, Japan in 1984, 1986 and 1996, respectively. His research interests include algorithmic learning theory, knowledge discovery and machine learning.

**Naoto Taketa** received the B.S. and M.S. degrees in Information Sciences from Hiroshima City University, in 2022 and 2024, respectively. He is currently with Rakuten Card Co., Ltd., System Strategy Department, System Division.