Proceedings of Machine Learning Research 76:1–37, 2017 Algorithmic Learning Theory 2017

Erasing Pattern Languages Distinguishable by a Finite Number of Strings*

Fahimeh Bayeh BAYEH22F@UREGINA.CA
Ziyuan Gao GAO257@CS.UREGINA.CA
Sandra Zilles ZILLES@CS.UREGINA.CA
Department of Computer Science, University of Regina, Regina, SK, Canada

Editor: Steve Hanneke and Lev Reyzin

Abstract

Pattern languages have been an object of study in various subfields of computer science for decades. This paper introduces and studies a decision problem on patterns called the *finite distinguishability* problem: given a pattern π, are there finite sets T^+ and T^- of strings such that the only pattern language containing all strings in T^+ and none of the strings in T^- is the language generated by π? This problem is related to the complexity of teacher-directed learning, as studied in computational learning theory, as well as to the long-standing open question whether the equivalence of two patterns is decidable. We show that finite distinguishability is decidable if the underlying alphabet is of size other than 2 or 3, and provide a number of related results, such as (i) partial solutions for alphabet sizes 2 and 3, and (ii) decidability proofs for variants of the problem for special subclasses of patterns, namely, regular, 1-variable, and non-cross patterns. For the same subclasses, we further determine the values of two complexity parameters in teacher-directed learning, namely the *teaching dimension* and the *recursive teaching dimension*.

Keywords: pattern languages, teaching dimension, recursive teaching dimension

1. Introduction

Database theory, pattern matching, computational learning theory, formal language theory—in these and other subfields of computer science a set L of strings is often represented by some string expression that "matches" all the strings in L. Regular expressions (as well as variants of regular expressions) are perhaps the most prominent such type of expression, but another kind of expression of relevance to many applications is the *pattern*. A pattern π is a finite string of constant symbols (often called terminal symbols) and variables, where the constant symbols are taken from some alphabet Σ. A string w over Σ matches π (or π matches w) if w can be obtained by substituting the variables in π with finite strings over Σ; the language of π, denoted $L(\pi)$, is then the set of all strings matching π. Angluin's original definition of pattern languages (Angluin, 1980) required that no variable be *erased*, i.e., substituted by the empty string, when matching a string; the corresponding pattern languages are hence called non-erasing pattern languages. In this paper, we study the case of so-called erasing or extended pattern languages (Shinohara, 1982b), where substitutions with the empty string are allowed. For example, the pattern $ax_1x_1abx_2$ over $\Sigma = \{a, b\}$ matches all strings starting with the symbol a, followed by a (possibly empty) square and the string ab, and ending in

* This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Discovery Grants program and Canada Research Chairs program.

any (possibly empty) suffix. Unless stated otherwise, we will use the term "pattern languages" to refer to erasing pattern languages.

Several fundamental problems on pattern languages have been addressed in the literature pertinent to learning theory, as they are of relevance to the design of learning methods that identify pattern languages from examples or from queries. For instance, the *membership problem*, i.e., to decide whether a given pattern matches a given string, is NP-complete (Jiang et al., 1994), and only a few interesting special cases are known in which it has a polynomial-time solution (Fernau and Schmid, 2015). Worse yet, the *inclusion problem*, to decide whether one given pattern generates a language contained in that of another, is undecidable (Freydenberger and Reidenbach, 2010). A prominent open question concerns the problem to decide whether two given patterns generate the same language, known as the *equivalence problem*. To date, it is not known whether this problem is decidable; notable decidable special cases were published around 20 years ago (Jiang et al., 1994; Ohlebusch and Ukkonen, 1997), but rather limited progress has been made on this problem since, cf. (Freydenberger and Reidenbach, 2010) for a discussion.

The focus of this paper is on the following decision problem, which we call the *finite distinguishability problem*: is a given pattern π finitely distinguishable (w.r.t. the class of all patterns), i.e., are there finite sets T^+ and T^- of strings such that $L(\pi)$ is the only pattern language that contains all of the strings in T^+ and none of the strings in T^-? This problem is of relevance to computational learning theory as well as to formal language theory; previously it has been studied in computational biology (Brazma et al., 2009) and in a recursion-theoretic context (Beros et al., 2016). For the non-erasing case, the problem is trivial since every pattern is finitely distinguishable w.r.t. the class of all patterns (Angluin, 1980). As it turns out, the erasing case is more complex.

In computational learning theory, finite distinguishability is equal to the property that $L(\pi)$ has a finite *teaching set* w.r.t. the class of all pattern languages. A teaching set T for a language L w.r.t. a class $\mathcal{L}$ containing L is a set of strings, each labeled either $+$ or $-$, such that L is the only language in $\mathcal{L}$ that contains all the $+$-labeled and none of the $-$-labeled strings in T. The size of a smallest teaching set is a lower bound on the number of labeled strings a learning algorithm would require to exactly identify L within $\mathcal{L}$ (Goldman and Kearns, 1995; Shinohara and Miyano, 1991).

From a language-theoretic point of view, the finite distinguishability problem is interesting in its own right, since the structure of teaching sets reveals structural properties of language classes. In the context of pattern languages in particular, there is another potential benefit of studying the finite distinguishability problem, due to its relevance to the unsolved equivalence problem. Firstly, if a pattern π is finitely distinguishable as witnessed by sets T^+ and T^- that can be algorithmically derived from π , then the problem of equivalence of π to any other pattern π' is decidable: it suffices to test whether π' matches all strings in T^+ and no strings in T^-. Secondly, if neither of two patterns π, π' is finitely distinguishable, then we know that a procedure deciding the equivalence problem on the instance (π, π') cannot solely rely on membership testing using the entire teaching set of either π or π'.

Our contributions are as follows: (i) We show that the finite distinguishability problem is decidable for all alphabet sizes other than 2 and 3. In doing so, we reveal some connections to the problem of deciding whether a pattern generates a regular language, which has previously been proven decidable for alphabet sizes other than 2 and 3 (Jain et al., 2010). (ii) For alphabet sizes 2 and 3, we provide partial results, again aligning with the existing literature on regular languages generated by patterns (Reidenbach and Schmid, 2014). (iii) We study variants of the finite distinguishability problem, namely, the question whether a pattern in class Π is finitely distinguishable from all patterns in class Π, for subclasses Π of the class of all patterns over a fixed alphabet. It turns out that this problem is decidable for the well-known classes of regular patterns, 1-variable patterns, and non-cross patterns.[1] Furthermore, for each of these classes, we prove that any finitely distinguishable pattern π has a teaching set of size polynomial in the length of π (linear for regular patterns, cubic for 1-variable patterns, while for non-cross patterns there is only one pattern, up to equivalence, with finite distinguishability). (iv) Due to the links to computational learning theory, we further explore the worst-case complexity of teaching pattern languages in two popular models of computational teaching, namely the teaching dimension model (Goldman and Kearns, 1995; Shinohara and Miyano, 1991) and the recursive teaching dimension model (Zilles et al., 2011), thus complementing an earlier such study on non-erasing pattern languages (Gao et al., 2016).

All our proofs establishing the finite distinguishability of some form of patterns are constructive in that they provide finite teaching sets rather than just proving their existence. They are thus meaningful for the design of strategies for algorithmic teaching and learning.

2. Preliminaries

$\mathbb{N}_0$ denotes the set of natural numbers $\{0, 1, 2, \ldots\}$ and $\mathbb{N} = \mathbb{N}_0 \setminus \{0\}$. For any set A, $|A|$ denotes the cardinality of A. If $a, b \in \mathbb{N}_0$ and $a \leq b$, $[a, b]$ denotes the interval $\{x \in \mathbb{N}_0 : a \leq x \leq b\}$. Let $X = \{x_1, x_2, x_3, \ldots\}$ be an infinite set of variable symbols. An alphabet is a finite or countably infinite set of symbols, disjoint from X. Given an alphabet Σ, a pattern is a non-empty finite string over $\Sigma \cup X$. The language $L(\pi)$ generated by a pattern π over Σ consists of all strings generated from π when replacing variables in π with any string over Σ, where all occurrences of a single variable must be replaced by the same string. For example, if $\pi = x_1 x_2 a b x_2$ and $\Sigma = \{a, b\}$, then $L(\pi)$ contains the strings ab, aab, $abbabbb$, but not the string $aabb$. This type of language is usually called extended or erasing pattern language (Shinohara, 1982b), to distinguish it from Angluin's notion of pattern language, which does not allow for a variable to be replaced by the empty string (Angluin, 1980). Patterns π and τ over Σ are said to be *equivalent* iff $L(\pi) = L(\tau)$. We often omit any reference to Σ when the choice of alphabet is clear from the context.

For any alphabets A and B, a *morphism* is a function $h : A^* \to B^*$ with $h(uv) = h(u)h(v)$ for all $u, v \in A^*$. A *substitution* (or *assignment*) is a morphism $h : (\Sigma \cup X)^* \to \Sigma^*$ with $h(a) = a$ for all $a \in \Sigma$. Given strings $w_1, \ldots, w_k \in (\Sigma \cup X)^*$ and a pattern $\pi \in (\Sigma \cup X)^*$ containing variables $x_1, \ldots, x_k$, $\pi[x_1 \to w_1, \ldots, x_k \to w_k]$ denotes the string derived from π by substituting w_i for x_i whenever $i \in [1, k]$.

For any set Γ of symbols, $\Gamma^+ = \Gamma^* \setminus \{\varepsilon\}$ is the set of non-empty words over Γ. For $w \in \Gamma^+$, $|w|$ denotes the length of w. For any $p \in [1, \ldots, |w|]$, $w[p]$ is the p^{th} symbol of w. For a symbol a and any $n \in \mathbb{N}_0$, a^n denotes the string equal to n concatenated copies of a. (Thus, a^0 is the empty string.)

Let Π^z denote the class of patterns over some specific alphabet Σ such that $|\Sigma| = z$. For any $\pi \in \Pi^z$, let $\text{Var}(\pi)$ denote the set of all distinct variables occurring in π, $\text{Const}(\pi)$ denote the set of all constant symbols occurring in π, and let $\pi(\varepsilon)$ denote the string obtained from π by substituting the empty string for all variables in π. Similarly, if a is any symbol, $\pi(a)$ denotes the string obtained when substituting the symbol a for all variables in π. We will often assume that a pattern $\pi \in \Pi^z$ is *normalised* in the sense that the k variables occurring in π are named $x_1, \ldots, x_k$ in order of their first occurrences from left to right (or x if $k = 1$).

1. See Section 4 for a definition of these pattern classes.

Any pattern in Σ^+ is a *constant pattern*; those in X^+ are called *constant-free*. $\Pi^z_{cf} \subseteq \Pi^z$ denotes the subclass of constant-free patterns. A *regular* pattern contains no variable more than once; by "regular pattern languages" we refer to languages generated by regular patterns.

Let Σ be any alphabet. A labelled example is a pair (w, ℓ), where $w \in \Sigma^*$ and $\ell \in \{+, -\}$. If $\ell = +$, the example is called a positive example, otherwise it is called a negative example. Given any set T of labelled examples, let T^+ denote the set of positively labelled strings in T and let T^- denote the set of negatively labelled strings in T. A pattern π is consistent with T (or T is consistent with π) if $T^+ \subseteq L(\pi)$ and $T^- \subseteq (\Sigma^* \setminus L(\pi))$.

This paper is concerned with a decision problem we call the *finite distinguishability problem*: given an alphabet Σ, a pattern π, and a reference class Π of patterns, is there a finite set T such that π is the only pattern up to equivalence in Π that is consistent with T? If yes, we call π finitely distinguishable w.r.t. Π. In the terminology of computational learning theory, one would rephrase the question as whether π has a finite *teaching dimension* w.r.t. Π. The teaching dimension of π w.r.t. Π, denoted by $\mathrm{TD}(\pi, \Pi)$ is defined as $\mathrm{TD}(\pi, \Pi) = \min\{|T| \mid T \text{ is a teaching set for } \pi \text{ w.r.t. } \Pi\}$. A teaching set for π w.r.t. Π is a set T that is consistent with π, but with no other pattern in Π (up to equivalence). This notion was originally defined in the more general context of concept learning (Goldman and Kearns, 1995; Goldman and Mathias, 1996; Shinohara and Miyano, 1991).

3. Pattern Languages with Finite Teaching Dimension

In this section, we investigate the structural properties of patterns that are finitely distinguishable. We first give some preparatory definitions.

Definition 1 *Fix any alphabet Σ of size $z \leq \infty$. For any $\pi \in \Pi^z$ with $\pi = X_1 c_1 X_2 \ldots c_{n-1}\, X_n$, $X_1, \ldots, X_n \in X^*$ and $c_1, \ldots, c_{n-1} \in \Sigma^+$, call each nonempty block X_i a* maximal variable block *of π. Call a set $\{Y_1, \ldots, Y_k\}$ of maximal variable blocks of π* independent with respect to π *iff every variable x in some block Y_i does not occur in any maximal variable block $Z \notin \{Y_1, \ldots, Y_k\}$ of π. In particular, the set $\{Z_1, \ldots, Z_l\}$ of all maximal variable blocks of π is independent w.r.t. π. Call a variable x* free *w.r.t. π iff x occurs in π exactly once. A pattern π is called* block-regular *if each of its maximal variable blocks contains a free variable w.r.t. π (Jain et al., 2010).*

Jain et al. (2010) showed that any block-regular pattern π is equivalent to the pattern obtained from π by dropping all the variables that occur at least twice in π.

Theorem 2 *(Jain et al., 2010, Theorem 6(b)) Fix an alphabet Σ, and let $\pi = c_1 X_1 c_2 X_2 \ldots X_{n-1} c_n$ be a block-regular pattern, where $c_1, c_n \in \Sigma^*$, $X_1, \ldots, X_n \in X^+$ and $c_1, \ldots, c_{n-1} \in \Sigma^+$. Then π is equivalent to the regular pattern $\pi' = c_1 x_1 c_2 x_2 \ldots x_{n-1} c_n$.*

We now present the main result of this paper. It states that for $z = 1$ and $z \geq 4$, finite distinguishability is decidable. For $z \in \{2, 3\}$, it shows that the finite distinguishability problem is decidable when restricted to constant-free patterns.

Theorem 3 *Let $\pi \in \Pi^z$.*

1. *Suppose $z = 1$. Let $x_1, \ldots, x_l$ be all the distinct variables occurring in π. For all $i \in [1, l]$, let p_i denote the number of times that x_i occurs in π. Then π is finitely distinguishable w.r.t. Π^z iff $l \geq 1$ and $\gcd(p_1, \ldots, p_l) = 1$.*

2. *Suppose $z \geq 2$. If $\pi \in \Pi^z_{cf}$, then π is finitely distinguishable w.r.t. Π^z iff π contains some variable exactly once.*

3. *Suppose $z \geq 4$. Then π is finitely distinguishable w.r.t. Π^z iff the following conditions are satisfied:*

   (a) π is block-regular;

   (b) π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$;

   (c) π starts and ends with variables.

   In particular, π is finitely distinguishable w.r.t. Π^z iff π is equivalent to a pattern π' of the shape $y_1 a_1 y_2 a_2 \ldots a_k\, y_{k+1}$, where $k \geq 0$, $a_1, a_2, \ldots, a_k \in \Sigma$ and $y_1, y_2, \ldots, y_{k+1}$ are $k+1$ distinct variables.

Thus, if $z = 1$ or $z \geq 4$, there is a polynomial-time decider for the set $\{\pi \in \Pi^z : TD(\pi, \Pi^z) < \infty\}$. Furthermore, if $z \geq 2$, there is a polynomial-time decider for the set $\{\pi \in \Pi^z_{cf} : TD(\pi, \Pi^z) < \infty\}$.

Proof (Sketch) *Proof of (1).* (1) follows from generalised forms of Colloraries 9 and 10 in (Gao et al., 2015); further details are given in Appendix A.

Proof of (2). Suppose that $z \geq 2$. Fix any distinct $a, b \in \Sigma$. If π contains some variable exactly once, then $L(\pi) = L(x_1)$, so that $\{(a,+),(b,+)\}$ is a teaching set for π w.r.t. Π^z. If π contains no variable and T is a finite set of examples labelled consistently with π, then $\pi' = \pi x_1^m$ is consistent with T, where $m > \max\{|\alpha| : \alpha \in T^+ \cup T^-\}$; i.e., $\mathrm{TD}(\pi, \Pi^z) = \infty$. Now suppose that π contains at least one variable and every variable occurring in π appears in π at least twice. Assume towards a contradiction that π has a finite teaching set T w.r.t. Π^z. Choose $m > \max(\{|\alpha| : \alpha \in T^+ \cup T^-\} \cup \{|\pi|\})$. Consider the string

$$\beta = \underbrace{a^m b^m a^m}\underbrace{a^{m+1} b^{m+1} a^{m+1}} \ldots \underbrace{a^{2m} b^{2m} a^{2m}},$$

which is a concatenation of the strings $a^{m+i} b^{m+i} a^{m+i}$ for i increasing from 0 to m. We will show that for some appropriately chosen block Y of variables,

(I) $\beta\pi(\varepsilon) \in L(Y\pi) \setminus L(\pi)$;

(II) $L(Y\pi) \supseteq L(\pi)$;

(III) $w \in L(Y\pi) \setminus L(\pi)$ implies $|w| \geq m$.

Notice that items (I), (II) and (III) together imply that $Y\pi$ is consistent with T while $L(Y\pi) \neq L(\pi)$, which contradicts the fact that T is a teaching set for π w.r.t. Π^z. We first prove that $\beta\pi(\varepsilon) \notin L(\pi)$. Assume otherwise. Fix a substitution $A : (X \cup \Sigma)^* \mapsto \Sigma^*$ witnessing $\beta\pi(\varepsilon) \in L(\pi)$. Given any strings $\alpha \in \Sigma^*$ and $\rho \in (X \cup \Sigma)^*$, say that ρ *covers* α *w.r.t.* A iff α is a prefix of $A(\rho)$. Our method of proof is to show by induction that for all $i \in \{-1, \ldots, m\}$ (where β_{-1} is defined to be ε), the shortest prefix ρ_i of π that covers

$$\beta_i = a^m b^m a^m \ldots a^{m+i} b^{m+i} a^{m+i}$$

w.r.t. A satisfies $|\rho_i| \geq i + 1$. For $i = m$, this will imply that $|\pi| \geq |\rho_m| \geq m + 1$, a contradiction. There is nothing to prove for $i = -1$ since $\beta_{-1} = \varepsilon$. Now suppose the statement to be proven holds for $n = k$, that is, if ρ_k is the shortest prefix of π that covers $\beta_k = a^m b^m a^m \ldots a^{m+k} b^{m+k} a^{m+k}$, then $|\rho_k| \geq k+1$. Consider $\beta_{k+1} = a^m b^m a^m \ldots a^{m+k} b^{m+k} a^{m+k} a^{m+k+1} b^{m+k+1} a^{m+k+1}$. Let s be the last symbol of ρ_k; note that s is a variable (as π is constant-free). Suppose the string $\beta_{k+1} = \beta_k a^{m+k+1}\, b^{m+k+1} a^{m+k+1}$ is covered by ρ_k w.r.t. A. Then, since no proper prefix of ρ_k covers β_k and s occurs in π at least twice, $A(\pi)$ must contain at least two copies of the string $a^{m+k+1} b^{m+k+1}\, a^{m+k+1}$, which is impossible. Hence there is a nonempty string θ for which the shortest prefix of π covering β_{k+1} w.r.t. A is equal to $\rho_k\theta$, so that by the induction hypothesis, $|\rho_{k+1}| \geq k+2$. This proves $\beta\pi(\varepsilon) \notin L(\pi)$. Now pick distinct variables y_1 and y_2 not occurring in π, and set

$$Y = \underbrace{y_1^m y_2^m y_1^m}\,\underbrace{y_1^{m+1} y_2^{m+2} y_1^{m+1}} \cdots \underbrace{y_1^{2m} y_2^{2m} y_1^{2m}}.$$

Observe that $\beta\pi(\varepsilon) \in L(Y\pi)$, proving (I). Further, (II) and (III) follow directly from the choice of m and Y. Thus T is not a teaching set for π w.r.t. Π^z, so that $\mathrm{TD}(\pi, \Pi^z) = \infty$.

Proof of (3). The proof that π is finitely distinguishable if it satisfies (a), (b) and (c) will be deferred to Appendix B, where it will be shown, more generally, that over any finite alphabet of size at least 2, Conditions (a), (b) and (c) together imply finite distinguishability.

It remains to show that if π does not satisfy either (a), (b) or (c), then $\mathrm{TD}(\pi, \Pi^z) = \infty$.

Case (i): π is not block-regular. Then one can fix some interval $[j_1, j_2]$ such that $\pi[j_1] \ldots \pi[j_2]$ is a maximal variable block of π and for all $j' \in [j_1, j_2]$, $\pi[j']$ occurs in π at least twice.

Suppose T were a finite teaching set for $L(\pi)$ w.r.t. Π^z. Choose $m > \max(\{|\alpha| : \alpha \in T^+ \cup T^-\} \cup \{|\pi|\})$, and let π' be the pattern obtained from π by inserting

$$Y = \underbrace{y_1^m y_2^m y_1^m}\,\underbrace{y_1^{m+1} y_2^{m+1} y_1^{m+1}} \cdots \underbrace{y_1^{2m} y_2^{2m} y_1^{2m}},$$

which is a concatenation of $y_1^{m+i} y_2^{m+i} y_1^{m+i}$ for i increasing from 0 to m, into π just before the j_1^{th} symbol of π, where $y_1, y_2 \notin \mathrm{Var}(\pi)$ are distinct variables. Choose distinct $d_1, d_2 \in \Sigma$ that are different from the last constant before the j_1^{th} symbol of π (suppose this occurs at the p_1^{th} position of π; $p_1 = 0$ if no such constant exists) and the first constant after the j_2^{th} symbol of π (suppose this occurs at the p_2^{th} position of π; $p_2 = |\pi| + 1$ if no such constant exists). Such d_1 and d_2 exist because $|\Sigma| \geq 4$. Let β be the string obtained from Y by substituting d_1 for y_1 and d_2 for y_2. Let γ be the string obtained from π by substituting d_1 for y_1, d_2 for y_2, and ε for every $x \in \mathrm{Var}(\pi)$. Then γ is of the form $C_1\beta C_2$, where $C_1C_2 \in \Sigma^*$ is the constant part of π. We claim that $\gamma \notin L(\pi)$. Suppose otherwise, and that $A'' : (X \cup \Sigma)^* \mapsto \Sigma^*$ witnesses $\gamma \in L(\pi)$.

Case (i.1): π contains at least one constant and $C_1 \neq \varepsilon$. Suppose

$$\gamma = \underbrace{a_1} \cdots \underbrace{a_i}\,\underbrace{\beta}\,\underbrace{a_{i+1}} \cdots \underbrace{a_l}, \qquad (1)$$

where $a_j \in \Sigma \cup \{\varepsilon\}$ for $j \in [1, l]$ and $a_i \in \Sigma$; note that $C_1 = a_1 \ldots a_i$ and $C_2 = a_{i+1} \ldots a_l$. A'' induces a mapping $I_{A''}$ from the set of all intervals of positions of π to the set of all intervals of positions of γ such that if $[p'_1, p'_2]$ and $[p'_2, p'_3]$ are mapped to $[q'_1, q'_2]$ and $[q'_3, q'_4]$ respectively, then $I_{A''}$ maps $[p'_1, p'_3]$ to $[q'_1, q'_4]$. Since it is a bit more convenient to speak of mappings from a specific occurrence of a subpattern of π to a specific occurrence of a substring of γ, we shall fix the convention that for any subpattern $\pi'' = \pi[p'_1] \ldots \pi[p'_\ell]$ of π and any $\alpha \in \{a_j : 1 \leq j \leq l\} \cup \{\beta\}$, "$I''_{A''}$ maps π'' to α" means that $I_{A''}$ maps $[p'_1, p'_\ell]$ to the interval of positions corresponding to the specific occurrence of α in γ indicated by braces in the decomposition (1).

If $I_{A''}$ maps the p_1^{th} symbol of π to some a_h with $h < i$, then it must also map the second to last constant symbol before the j_1^{th} symbol of π to some $a_{h'}$ with $h' < h$; applying this argument successively then leads to a contradiction. A similar argument shows that $I_{A''}$ cannot map the p_1^{th} symbol of π to some a_h with $h > i$. Furthermore, by the choice of d_1 and d_2, $I_{A''}$ cannot map its p_1^{th} position to any symbol in β. Hence $I_{A''}$ maps the p_1^{th} symbol of π to a_i. In particular, $I_{A''}$ maps the suffix of π starting from its $(p_1+1)^{st}$ symbol to the suffix βC_2 of γ. Since d_1 and d_2 are different from the constant symbol in π's p_2^{th} position, $I_{A''}$ maps the maximal variable block of variables $\pi[j_1]\ldots\pi[j_2]$ to β. Note that $I_{A''}$ cannot map $\pi[j_1]\ldots\pi[j_2]$ to any proper extension of β because otherwise γ (as reasoned above) would not be "long enough". By the choice of $[j_1, j_2]$, for every $j' \in [j_1, j_2]$, $\pi[j'] \in X$ and $\pi[j']$ occurs in π at least twice. Note that for every $j' \in [j_1, j_2]$ such that $I_{A''}(j') \neq \varepsilon$, $\pi[j']$ neither occurs before the j_1^{th} position of π nor occurs after the j_2^{th} position of π because otherwise the length of γ would have to increase by at least one. Hence the subpattern $\pi[j_1+i_1]\ldots\pi[j_1+i_h]$ of π that $I_{A''}$ maps to β such that $I_{A''}(j_1+i_j) \neq \varepsilon$ whenever $1 \leq j \leq h$ is of the shape $q_1^{n_1} q_2^{n_2} \ldots q_s^{n_s}$, where $q_1, q_2, \ldots, q_s \in X$ and each q_i occurs in $\pi[j_1+i_1]\ldots\pi[j_1+i_h]$ at least twice. But an argument similar to that in the proof of statement (2) above shows that β cannot match any such block Q of variables $q_1^{n_1} q_2^{n_2} \ldots q_s^{n_s}$, where each q_i occurs in Q at least twice and $|Q| < m$. Thus $\gamma \notin L(\pi)$, indeed.

Case (i.2): $C_1 = \varepsilon$ but $C_2 \neq \varepsilon$. This case can be argued similarly to Case 1.

Case (i.3): π is constant-free. Then π is of the shape $r_1^{n_1} r_2^{n_2} \ldots r_s^{n_s}$, where $r_1, r_2, \ldots, r_s \in X$ and (since π is not block-regular) each r_i occurs in π at least twice; hence an argument similar to that in the proof of statement (2) shows that $\gamma \notin L(\pi)$.

By construction, $\gamma \in L(\pi')$. As π' is consistent with T, $\mathrm{TD}(\pi, \Pi^z) = \infty$.

Case (ii): π contains a substring of the form ab, where $a, b \in \Sigma$. (a and b are not necessarily distinct.) Since $|\Sigma| \geq 4$, one can fix some $c \in \Sigma$ with $c \notin \{a, b\}$. Let j_3 be a position of π such that $\pi[j_3]\pi[j_3+1] = ab$. If $L(\pi)$ had a finite teaching set T w.r.t. Π^z, then one can argue as in Case (i) that there is a positive m so large that if π' is obtained from π by inserting y^m between the j_3^{th} and $(j_3+1)^{st}$ positions of π for some variable $y \notin \mathrm{Var}(\pi)$, then π' would be consistent with T. On the other hand, let γ be the string derived from π' by substituting c for y and ε for every other variable; note that the number of times the substring ab occurs in γ is strictly less than the number of times that ab occurs in π, which implies $\gamma \notin L(\pi)$ and so $L(\pi') \neq L(\pi)$. Therefore $\mathrm{TD}(\pi, \Pi^z) = \infty$.

Case (iii): π starts or ends with a constant symbol (or both). Suppose π starts with the constant symbol a. The proof that $L(\pi)$ has no finite teaching set w.r.t. Π^z is very similar to that in Case (ii); the only difference here is that one chooses some $b \in \Sigma \setminus \{a\}$ and considers $\pi' = y^m\pi$ for some variable $y \notin \mathrm{Var}(\pi)$ and a sufficiently large m. In this case, $b^m\pi(\varepsilon) \in L(\pi') \setminus L(\pi)$, and therefore $L(\pi') \neq L(\pi)$. An analogous argument holds if π ends with a constant symbol.

This completes the proof of the characterisation.

Finally, note that there are polynomial-time algorithms to (i) determine whether or not the greatest common divisor of a set of positive integers is equal to 1, (ii) determine whether or not a given pattern $\pi \in \Pi^z_{cf}$ contains a variable that occurs exactly once, and (iii) determine whether or not any given $\pi \in \Pi^z$ satisfies conditions (a), (b) and (c) in statement (3). For (iii), note that π is block-regular iff every maximal block Y of π contains a free variable, and this condition can be checked in $O(|\pi|)$ steps. Further, it takes $O(|\pi|)$ steps to check whether or not π contains a substring $\alpha \in \Sigma^+$ such that $|\alpha| = 2$ and another $O(|\pi|)$ steps to determine whether or not π starts and ends with variables. Thus for any given $z \geq 2$, the set $\{\pi \in \Pi^z_{cf} : \text{TD}(\pi, \Pi^z) < \infty\}$ has a polynomial-time decider; similarly, for $z \notin \{2, 3\}$, the set $\{\pi \in \Pi^z : \text{TD}(\pi, \Pi^z) < \infty\}$ is polynomial-time decidable. ■

In fact, the conditions in Theorem 3(3) are sufficient for any pattern over an alphabet of size at least 2 to be finitely distinguishable. We prove this in Appendix B.

Proposition 4 *Let $\pi \in \Pi^z$ and $z \geq 2$. Then π is finitely distinguishable w.r.t. Π^z if π is equivalent to a pattern of the shape $y_1 a_1 y_2 \ldots a_k y_{k+1}$, where $a_1, \ldots, a_k \in \Sigma$ and $y_1, \ldots, y_{k+1}$ are distinct variables.*

Jain et al. (2010) showed that for every pattern π over any finite alphabet with at least 4 letters, $L(\pi)$ is a regular language iff π is block-regular. This yields the following corollary.

Corollary 5 *Suppose $4 \leq z < \infty$ and $\pi \in \Pi^z$. Then π is finitely distinguishable w.r.t. Π^z iff all of the following conditions are satisfied:*

1. *$L(\pi)$ is regular;*
2. *π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$;*
3. *π starts and ends with variables.*

The remaining part of this section is devoted to the question of whether Theorem 3(3) (or some slight variation) extends to alphabets Σ with $|\Sigma| \in \{2, 3\}$. We shall illustrate with examples the failure of Theorem 3(3) for alphabets that have exactly two or three letters. In particular, over such alphabets, it will be seen that the structure of finitely distinguishable patterns can be fairly complex, which suggests that the problem of deciding finite distinguishability of π w.r.t. Π^z for $z \in \{2, 3\}$ and any $\pi \in \Pi^z$ may be more difficult than for the case $z \geq 4$.

Example 1 *Let $\Sigma = \{a, b\}$ and $\pi = x_1 a x_2^2 b x_3$. Note that π is not block-regular. Let $\pi' = x_1 ab x_2$. We claim that $L(\pi') = L(\pi)$. $L(\pi') \subseteq L(\pi)$ is immediate. Consider any $\beta \in \Sigma^*$ obtained from π by substituting α_i for x_i, where $i \in \{1, 2, 3\}$. Since $a\alpha_2^2 b$ must contain the substring ab, β is of the shape $\gamma_1 ab \gamma_2$, where $\gamma_1, \gamma_2 \in \Sigma^*$, and so $\beta \in L(\pi')$. Therefore $L(\pi') = L(\pi)$. Furthermore, observe that $\{(ab, +), (a, -), (b, -), (baba, +)\}$ is a (finite) teaching set for π' w.r.t. Π^2. Thus the characterisation obtained in Theorem 3(3) does not apply to alphabets with exactly two letters.*

The next example shows that Theorem 3(3) does not apply to the class of erasing pattern languages over any alphabet of size 3. The corresponding proof is given in Appendix C.

Example 2 *Let $\Sigma = \{a, b, c\}$ and $\pi = x_1 x_2 x_3 a x_2 x_4^2 x_5^3 x_6 b x_7 x_6 x_8$. Then π is finitely distinguishable w.r.t. Π^3 but $L(\pi)$ cannot be generated by any regular pattern. Note that while π is not regular, it generates a regular language, namely, $L(\pi) = \Sigma^* ab \Sigma^* \cup \Sigma^* c \Sigma^* acb \Sigma^* \cup \Sigma^* acb \Sigma^* c \Sigma^* \cup \Sigma^* ac^2 c^* b \Sigma^*$.*

The next example shows that over any alphabet of size exactly 2, there is a pattern π that is finitely distinguishable w.r.t. Π^2 while $L(\pi)$ cannot be generated by any regular pattern.

Example 3 *Let $\Sigma = \{a, b\}$ and $\pi = x_1 x_2 a x_2 x_3^2 x_4^3 x_5 a x_5 x_6$. Then π is finitely distinguishable w.r.t. Π^2 but $L(\pi)$ cannot be generated by any regular pattern.*

Proof (Sketch.) One may show that $\{(aa,+),(a,-),(baa,+),(aab,+),\ (ab^2a,+),(ab^3a,+),(aba,-),(abab,+),(ababa,+),(baba,+)\}$ is a teaching set for π w.r.t. Π^2. Furthermore, if $L(\pi)$ were generated by some regular pattern τ, then τ must be of the shape $x_1ax_2ax_3$. But $aba \in L(x_1ax_2ax_3) \setminus L(\pi)$, and so $L(\pi) \neq L(\tau)$. ■

Theorem 3(1), Example 2 and Example 3 together imply that one direction of the characterisation in Theorem 3(3) – that $\mathrm{TD}(\pi,\Pi^z) < \infty \Rightarrow \pi$ satisfies Conditions (a), (b) and (c) – applies only to the case $z \geq 4$. The next two examples from (Reidenbach and Schmid, 2014) and (Jain et al., 2010) show that the reverse direction of Theorem 3(3) fails for $z \in \{2,3\}$ as well if one relaxes Condition (a) by only requiring that $L(\pi)$ be a regular language. Their proofs are in Appendices E and F, resp.

Example 4 *Let $\Sigma = \{a,b\}$ and $\pi = x_1ax_2^2ax_3$. Then (a) $L(\pi)$ is regular, (b) π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$, and (c) π starts and ends with variables, but π is not finitely distinguishable w.r.t. Π^2.*

Example 5 *Let $\Sigma = \{a,b,c\}$ and $\pi = x_1x_2x_3ax_2x_4^2x_5bx_6x_5x_7$. Then (a) $L(\pi)$ is regular, (b) π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$, and (c) π starts and ends with variables, but π is not finitely distinguishable w.r.t. Π^3.*

According to Examples 2 and 3, a pattern language over any alphabet of size 2 or 3 may be finitely distinguishable without being generable by a block-regular pattern. Our next result shows, on the other hand, that over *any* finite alphabet, a finitely distinguishable pattern language must necessarily be regular. The converse of the latter statement (even with restrictions on the length of every constant block of the pattern and on the first as well as last symbols of the pattern) is false, as we have seen in Examples 4 and 5. The proof of Theorem 6 is given in Appendix G.

Theorem 6 *Let $1 \leq z < \infty$ and $\pi \in \Pi^z$. If π is finitely distinguishable w.r.t. Π^z, then $L(\pi)$ is regular.*

The following theorem provides necessary conditions for a pattern to be finitely distinguishable w.r.t. the whole class of patterns over any alphabet of size 2 or 3. It is proven in Appendix H.

Theorem 7 *Let $z \in \{2,3\}$, $\Sigma_1 = \{a,b\}$, $\Sigma_2 = \{a,b,c\}$ and $\pi = X_1c_1X_2c_2\ldots X_{n-1}c_{n-1}X_n$, where $X_2,\ldots,X_{n-1} \in X^+$, $c_1,\ldots,c_{n-1} \in \Sigma_1^+$ if $z = 2$, $c_1,\ldots,c_{n-1} \in \Sigma_2^+$ if $z = 3$, and $X_1, X_n \in X^*$. If π is finitely distinguishable w.r.t. Π^z, then the following conditions hold for all $i \in [1,n-1]$.*

1. *If $z = 2$, then $c_i \in \{a,b,ab,ba\}$; if $z = 3$, then $c_i \in \Sigma_2$.*
2. *If $z = 2$, then for all $\alpha \in \{X_1, X_n, \delta X_i\delta, \delta X_i\overline{\delta}X_{i+1}\delta, \delta\overline{\delta}X_i\delta, \delta X_i\overline{\delta}\delta\}$ such that α is a substring of π, where $\delta,\overline{\delta} \in \Sigma$ and $\delta \neq \overline{\delta}$, there is a $k \geq 1$ for which α contains variables $y_1,\ldots,y_k$ such that for all $j \in [1,k]$, y_j occurs q_j times in α for some $q_j \geq 1$, y_j does not occur outside the block α and $\gcd(q_1,\ldots,q_k) = 1$. If $z = 3$, then the latter statement holds for $\alpha = X_i$.*
3. *If $z = 2$, then π contains at least one free variable; if $z = 3$, then X_1 and X_n each contains at least one free variable.*

4. Interesting Subclasses of Pattern Languages

This section presents some results on various subclasses of the class of all pattern languages, namely the classes of (i) regular pattern languages, (ii) 1-variable pattern languages, and (iii) non-cross pattern languages. These have previously been studied in the literature on erasing pattern languages (Erlebach et al., 2001; Shinohara, 1982b; Reidenbach, 2006), because certain decision problems or learning problems that are infeasible or unsolvable in the general case have simple solutions for these subclasses.

A 1-variable pattern is a pattern that contains at most 1 variable (possibly with repetitions), while a non-cross pattern[2] is of the shape $x_1^{i_1} x_2^{i_2} \cdots x_n^{i_n}$ for some $n \geq 1$. Pattern languages generated by 1-variable patterns (non-cross patterns, resp.) are called 1-variable pattern languages (non-cross pattern languages, resp.). The class of all 1-variable (non-cross, regular, resp.) patterns over an alphabet of size z is denoted by $1\Pi^z$ ($\mathrm{NC}\Pi^z$, $\mathrm{R}\Pi^z$, resp.).

Characterizing finitely distinguishable patterns turns out somewhat simpler for these special cases. In particular, finite distinguishability w.r.t. any such reference class is decidable.

The class of regular erasing pattern languages is learnable with polynomially many membership queries (i.e., questions of the kind "does the string w match the unknown pattern?") iff the learner is initially given a string from the target language (Nessel and Lange, 2005). Note that the membership query complexity is also an upper bound on the teaching dimension. The next theorem, which is proven in Appendix I, gives a linear upper bound on $\mathrm{TD}(\pi, \mathrm{R}\Pi^z)$ for any regular pattern π.

Theorem 8 *Let $z \in \mathbb{N} \cup \{\infty\}$ and let π be a regular pattern over Σ. Then $TD(\pi, R\Pi^z) \leq 2|\pi| + 1$.*

The class $1\Pi^z$ of 1-variable patterns has been treated quite extensively in the literature. In particular, the corresponding class of *non-erasing* languages is efficiently learnable from queries (Erlebach et al., 2001) while its membership problem is decidable in polynomial time (Angluin, 1980). By contrast, the class of *erasing* 1-variable pattern languages is not learnable in various models of query learning (Nessel and Lange, 2005). Theorem 9 shows that the finite distinguishability problem restricted to $1\Pi^z$ has a simple decision procedure; further, any 1-variable pattern π with finite teaching dimension w.r.t. $1\Pi^z$ has a teaching set of size at most cubic in $|\pi|$.

Theorem 9 *Let $z \in \mathbb{N} \cup \{\infty\}$ and let π be a 1-variable pattern over Σ. Then $TD(\pi, 1\Pi^z) < \infty$ iff π contains a variable. If π contains a variable, then $TD(\pi, 1\Pi^z) = O(|\pi|)$ if $z = 1$ and $TD(\pi, 1\Pi^z) = O(|\pi|^3)$ if $z \geq 2$ (including $z = \infty$).*

Proof (Sketch) If π contains no variable and T is a finite set of examples labeled consistently with π, then $\pi' = \pi x_1^m$ is a 1-variable pattern consistent with T, where $m > \max\{|\alpha| : \alpha \in T^+ \cup T^-\}$. Consequently, $\mathrm{TD}(\pi, 1\Pi^z) = \infty$. If π contains a variable, pick $a \in \Sigma$. Then π is one of only finitely many 1-variable patterns consistent with the set $T = \{(\pi(\varepsilon), +), (\pi(a), +)\}$ and thus $\mathrm{TD}(\pi, 1\Pi^z) < \infty$. To see this, suppose π' is any 1-variable pattern consistent with T. Obviously, $|\pi(\varepsilon)|$ upper-bounds the number of constants in π' and the value $|\pi(a)| - |\pi(\varepsilon)|$, which is greater than zero by the choice of π, upper-bounds the number of variable positions in π'. Thus, there are only finitely many such π'. The rest of the proof is in Appendix J. ■

Non-cross patterns were introduced by Shinohara (1982a) as a form of pattern for which the membership problem is polynomial-time solvable, in contrast to the NP-completeness of the membership

2. In this paper, a "non-cross pattern" will always refer to a *constant-free* non-cross pattern.

problem for the general class of patterns (Angluin, 1980; Jiang et al., 1994). Non-cross erasing pattern languages are also learnable in the limit for any alphabet (Reidenbach, 2006). The finite distinguishability problem restricted to the class of all non-cross patterns turns out to be quite straightforward; in fact, over any alphabet Σ with $z = |\Sigma| \geq 2$, there is only one non-cross pattern (up to equivalence) with finite teaching dimension w.r.t. NCΠ^z.

Theorem 10 *Let $z \in \mathbb{N} \cup \{\infty\}$ and let $\pi = x_1^{n_1} \ldots x_k^{n_k}$ be a non-cross pattern over Σ.*

1. *Let $z = 1$. Then TD$(\pi, NC\Pi^z) < \infty$ iff the greatest common divisor of $n_1, \ldots, n_k$ is 1.*
2. *Let $z \geq 2$. Then TD$(\pi, NC\Pi^z) < \infty$ iff $n_i = 1$ for some $i \in [1, k]$, i.e., iff π contains at least one non-repeated variable.*

Proof (Sketch) Statement 1 was proven in (Gao et al., 2015, Corollaries 9 and 10). To prove 2, first suppose $n_i = 1$ for some $i \in [1, k]$. Then $L(\pi) = \Sigma^*$ and $\{(\varepsilon, +), (a, +)\}$ is a teaching set for π w.r.t. NCΠ^z. Next suppose $n_i \geq 2$ for all $i \in [1, k]$ and let T be a finite set of labeled examples consistent with π. Pick the first variable not occurring in π (say x_{k+1}) and define $\pi' = \pi x_{k+1}^{n_{k+1}}$ where $n_{k+1} > \max\{|\alpha| : \alpha \in T^+ \cup T^-\}$ and $n_{k+1} > |\pi|$. Note that $L(\pi) \subset L(\pi')$. Indeed, choose a sequence $m_1, \ldots, m_k, m_{k+1}$ such that $m_i n_i < m_{i+1} n_{i+1}$ for all $i \leq k$. Let a and b be two distinct letters in Σ and assume that π is normalised. Then the string obtained from π' by replacing every odd-indexed variable x_{2i-1} with $a^{m_{2i-1}}$ and every even-indexed variable x_{2i} with $b^{m_{2i}}$ is in $L(\pi') \setminus L(\pi)$ (for a formal proof, see Appendix K). π' cannot generate any of the negative examples in T, so that π' is a non-cross pattern consistent with T. We conclude that TD$(\pi, \text{NC}\Pi^z) = \infty$. ■

5. Worst-Case Teaching Complexity

In computational learning theory, the teaching dimension of a class of concepts refers to the worst-case number of examples a teacher needs to present to the learner in order to teach any concept in the class. If Π is any class of patterns, the teaching dimension of the class of languages generated by patterns in Π, denoted by TD(Π), is defined as TD$(\Pi) = \sup\{\text{TD}(\pi, \Pi) : \pi \in \Pi\}$. This parameter indicates how difficult it is to distinguish single languages in the class from all others. The value of TD(Π) is finite iff there is an upper bound on the number of strings needed for solving this task.

All proofs in this section will be relegated to the appendix.

Since, by Theorem 3, for any alphabet size there are patterns with an infinite teaching dimension with respect to the class of all (erasing) pattern languages, it is obvious that TD$(\Pi^z) = \infty$ for all $z \in \mathbb{N} \cup \{\infty\}$. The same holds for 1-variable pattern languages and for non-cross pattern languages, by Theorems 9 and 10, which yields the following theorem.

Theorem 11 *Let $z \in \mathbb{N} \cup \{\infty\}$. Then TD$(\Pi^z)$ = TD$(1\Pi^z)$ = TD$(NC\Pi^z) = \infty$.*

By contrast, for $z \geq 7$ as well as for $z = 1$, the corresponding class of regular pattern languages has a finite teaching dimension (whose exact value depends on z).

Theorem 12

1. *TD$(R\Pi^1) = 3$.*

2. *For all* $z \geq 2$ *(including* $z = \infty$*),* $TD(R\Pi^z) \geq 5$.

3. *For all* $z \geq 7$ *(including* $z = \infty$*),* $TD(R\Pi^z) = 5$.

The teaching dimension model is just one of several models of teacher-directed learning that has been studied in the literature. A related model that has attracted the attention of the learning theory community due to its connections to the VC dimension (Vapnik and Chervonenkis, 1971) (arguably the most important complexity parameter studied in statistical learning theory) and to sample compression (Floyd and Warmuth, 1995) is the *recursive teaching* model (Zilles et al., 2011). Recursive teaching can be conceived to proceed in (possibly infinitely many) stages: in the first stage, one teaches (some or all of) the concepts that have a small enough teaching dimension w.r.t. the whole concept class. One then removes those concepts from the class and proceeds recursively with the remaining concepts. We here formulate the definition specifically for pattern languages.

Definition 13 (Zilles et al. (2011); Gao et al. (2015, 2016, 2017a)) *Let* Π *be a class of patterns. A recursive teaching sequence for* Π *is a sequence* $\mathcal{S} = ((S_0, d_0), (S_1, d_1), \ldots)$*, where* $\bigcup_{i \in \mathbb{N}} S_i = \Pi$ *is a disjoint union and, for all* $i \in \mathbb{N}$ *and all* $\pi \in S_i$*, we have* $d_i < \infty$*, where*

$$d_i = \sup\{TD(\pi, \bigcup_{j \geq i} S_j) : \pi \in S_i\}\,.$$

A teaching set for $\pi \in S_i$ *w.r.t.* $\bigcup_{j \geq i} S_j$ *is then called a recursive teaching set for* π *w.r.t.* $\mathcal{S}$*. The order* $ord(\mathcal{S})$ *of* $\mathcal{S}$ *is defined by* $ord(\mathcal{S}) = \sup\{d_i \mid i \in \mathbb{N}\}$*. Finally, the recursive teaching dimension of* Π*, denoted by* $RTD(\Pi)$*, is the smallest order over all recursive teaching sequences for* Π*, i.e.,* $RTD(\Pi) = \min\{ord(\mathcal{S}) \mid \mathcal{S}$ *is a recursive teaching sequence for* $\Pi\}$*.*

For the classes of one-variable and of non-cross pattern languages, it turns out that recursive teaching is not a suitable model and does not improve on the negative results from Theorem 11 concerning the teaching dimension. Depending on the class and alphabet size, either the RTD is infinite or no recursive teaching sequence exists.

Theorem 14

1. *If* $z \in \mathbb{N} \cup \{\infty\}$*, then no recursive teaching sequence for* $1\Pi^z$ *exists.*

2. *If* $z \in \mathbb{N} \cup \{\infty\}$ *and* $z \geq 2$*, then no recursive teaching sequence for* $NC\Pi^z$ *exists.*

3. $RTD(NC\Pi^1) = \infty$.

For regular pattern languages, recursive teaching is provably more efficient than teaching according to the classical model, for alphabet sizes different from 2, as the next theorem shows. Determining $\mathrm{RTD}(\mathrm{R}\Pi^2)$ remains an open problem.

Theorem 15 *Let* $z \in \mathbb{N} \cup \{\infty\}$*. If* $z \neq 2$*, then* $RTD(R\Pi^z) = 2$*.*

6. Conclusions

Finite distinguishability of patterns is a decision problem of relevance to computational learning theory and to the open question of whether the equivalence problem for erasing pattern languages is decidable. Since Ohlebusch and Ukkonen (1997) already proved decidability of the equivalence problem restricted to the types of patterns for which our paper proves finite distinguishability, our results do not directly yield new results on the equivalence problem. However, they establish that any equivalence test for two patterns failing our test for finite distinguishability must necessarily use more information than that provided solely by the membership of a finite set of strings.

Our study on the teaching dimension/recursive teaching dimension of classes of erasing pattern languages complements an earlier such study on non-erasing pattern languages (Gao et al., 2016).

We leave a number of open problems, most notably: (i) for alphabet sizes 2 and 3, characterize the patterns that are finitely distinguishable and determine whether finite distinguishability is decidable, (ii) determine $\mathrm{TD}(\mathrm{R}\Pi^z)$ for $2 \leq z \leq 6$, and (iii) determine $\mathrm{RTD}(\mathrm{R}\Pi^2)$. Recently, the new model of *preference-based teaching* was proposed, in particular to address cases of concept classes for which no recursive teaching sequence exists (Gao et al., 2017a). One can show that for alphabets of size at least 3, non-cross patterns can be taught in the preference-based model using just a single example (Gao et al., 2017b), while we have shown above that they do not possess a recursive teaching sequence. A detailed study of preference-based teaching of pattern languages may lead to further interesting insights into their structural properties.

Acknowledgements. We thank the referees of ALT 2017 for their helpful comments and suggestions; special thanks go to one referee for pointing out a flaw in an earlier proof of Theorem 10(2).

References

Dana Angluin. Finding patterns common to a set of strings. *J. Comput. Syst. Sci.*, 21:46–62, 1980.

Achilles Beros, Ziyuan Gao, and Sandra Zilles. Classifying the arithmetical complexity of teaching models. In *ALT*, pages 145–160, 2016.

Alvis Brazma, Inge Jonassen, Ingvar Eidhammer, and David Gilbert. Approaches to the automatic discovery of patterns in biosequences. *J. Comput. Biol.*, 5(2):279–305, 2009.

Thomas Erlebach, Peter Rossmanith, Hans Stadtherr, Angelika Steger, and Thomas Zeugmann. Learning one-variable pattern languages very efficiently on average, in parallel, and by asking queries. *Theor. Comput. Sci.*, 261(1):119–156, 2001.

Henning Fernau and Markus L. Schmid. Pattern matching with variables: A multivariate complexity analysis. *Inf. Comput.*, 242:287–305, 2015.

Sally Floyd and Manfred K. Warmuth. Sample compression, learnability, and the Vapnik-Chervonenkis dimension. *Machine Learning*, 21(3):269–304, 1995.

Dominik D. Freydenberger and Daniel Reidenbach. Bad news on decision problems for patterns. *Inf. Comput.*, 208:83–96, 2010.

Ziyuan Gao, Hans U. Simon, and Sandra Zilles. On the teaching complexity of linear sets. In *ALT*, pages 102–116, 2015.

Ziyuan Gao, Zeinab Mazadi, Regan Meloche, Hans U. Simon, and Sandra Zilles. Distinguishing pattern languages with membership examples. 2016. Accepted with minor revisions in *Information and Computation*.

Ziyuan Gao, Christoph Ries, Hans U. Simon, and Sandra Zilles. Preference-based teaching. *J. Mach. Learn. Res.*, 18:1–32, 2017a.

Ziyuan Gao, Christoph Ries, Hans U. Simon, and Sandra Zilles. Preference-based teaching of pattern languages. 2017b. Unpublished manuscript.

Sally A. Goldman and Michael J. Kearns. On the complexity of teaching. *J. Comput. Syst. Sci.*, 50: 20–31, 1995.

Sally A. Goldman and H. D. Mathias. Teaching a smarter learner. *J. Comput. Syst. Sci.*, 52(2): 255–267, 1996.

Sanjay Jain, Yuh Shin Ong, and Frank Stephan. Regular patterns, regular languages and context-free languages. *Inf. Process. Lett.*, 110:1114–1119, 2010.

Tao Jiang, Efim Kinber, Arto Salomaa, Kai Salomaa, and Sheng Yu. Pattern languages with and without erasing. *Int. J. Comput. Math.*, 50:147–163, 1994.

Tao Jiang, Arto Salomaa, Kai Salomaa, and Sheng Yu. Decision problems for patterns. *J. Comp. Sys. Sci.*, 50:53–63, 1995.

M. Lothaire. *Combinatorics on words*. Addison-Wesley Publishing Company, Inc., 1983.

Jochen Nessel and Steffen Lange. Learning erasing pattern languages with queries. *Theor. Comput. Sci.*, 348:41–57, 2005.

Enno Ohlebusch and Esko Ukkonen. On the equivalence problem for e-pattern languages. *Theor. Comput. Sci.*, 186(1–2):231–248, 1997.

Daniel Reidenbach. A non-learnable class of e-pattern languages. *Theor. ComputṠci.*, 350:91–102, 2006.

Daniel Reidenbach and Markus L. Schmid. Regular and context-free pattern languages over small alphabets. *Theor. Comput. Sci.*, 518:80–95, 2014.

Ayumi Shinohara and Satoru Miyano. Teachability in computational learning. *New Generation Comput.*, 8:337–347, 1991.

Takeshi Shinohara. Polynomial time inference of pattern languages and its application. In *7th IBM Symposium on Mathematical Foundation of Computer Science*, 1982a.

Takeshi Shinohara. Polynomial time inference of extended regular pattern languages. In *RIMS Symposium on Software Science and Engineering*, pages 115–127, 1982b.

Vladimir Vapnik and Alexey Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.

Sandra Zilles, Steffen Lange, Robert Holte, and Martin Zinkevich. Models of cooperative teaching and learning. *J. Mach. Learn. Res.*, 12:349–384, 2011.

Appendix A. Proof of Theorem 3(1)

Theorem 3(1). Let $\pi \in \Pi^1$. Let $x_1, \ldots, x_l$ be all the distinct variables occurring in π. For all $i \in [1, l]$, let p_i denote the number of times that x_i occurs in π. Then π is finitely distinguishable w.r.t. Π^1 iff $l \geq 1$ and $\gcd(p_1, \ldots, p_l) = 1$.

Proof It was shown in (Gao et al., 2015) (Corollaries 9 and 10) that the linear set $\{\boldsymbol{v}^\top \boldsymbol{x} : \boldsymbol{x} \in \mathbb{N}_0^n\}$ for any $n \geq 1$ and $\boldsymbol{v} = (v_1, \ldots, v_n) \in \mathbb{N}_0^n$ has finite teaching dimension w.r.t. the class $\{\{\boldsymbol{v}^\top \boldsymbol{x} : \boldsymbol{x} \in \mathbb{N}_0^n\} : \boldsymbol{v} \in \mathbb{N}_0^n \wedge n \geq 1\}$ iff $\gcd(v_1, \ldots, v_n) = 1$. Notice that for any $c \in \mathbb{N}_0^n$, $\{c + \boldsymbol{v}^\top \boldsymbol{x} : \boldsymbol{x} \in \mathbb{N}_0^n\}$ is the *commutative image* (or *Parikh image*) of the erasing pattern language generated by $a^c x_1^{v_1} x_2^{v_2} \ldots x_n^{v_n}$ over any unary alphabet $\{a\}$. Theorem 3(1) is thus a consequence of the following "shift lemma".

Lemma A.1 *Let $\mathcal{L}$ be a class of nonempty subsets of $\mathbb{N}_0$ such that $0 \in L$ for all $L \in \mathcal{L}$. Define the shift-extension $\mathcal{L}'$ of $\mathcal{L}$ by $\mathcal{L}' = \{c + L : (c \in \mathbb{N}_0) \wedge (L \in \mathcal{L})\}$. Then for all $c \in \mathbb{N}_0$ and $L \in \mathcal{L}$, TD$(L, \mathcal{L}) \leq$ TD$(c + L, \mathcal{L}') \leq c + 1 +$ TD$(L, \mathcal{L})$.*

Proof of Lemma A.1. We first prove $\mathrm{TD}(L, \mathcal{L}) \leq \mathrm{TD}(c + L, \mathcal{L}')$. Suppose for a contradiction that there exists a teaching set T for $c + L$ w.r.t. $\mathcal{L}'$ that has size smaller than $\mathrm{TD}(L, \mathcal{L})$. Define $T' = \{(x - c, +) : x \in T^+\} \cup \{(x - c, -) : x \in T^- \wedge x \geq c\}$. Note that T' is consistent with L. Since $|T'| < \mathrm{TD}(L, \mathcal{L})$, there exists some $L' \in \mathcal{L}$ such that L' is consistent with T' and $L' \neq L$. Consequently, $c + L'$ is consistent with $\{(c + y, +) : y \in T'^+\} \cup \{(c + y, -) : y \in T'^-\} \cup \{(x, -) : x \in T^- \wedge x < c\} = T$, a contradiction.

We next prove $\mathrm{TD}(c + L, \mathcal{L}') \leq c + 1 + \mathrm{TD}(L, \mathcal{L})$. Let T_1 be a teaching set for L w.r.t. $\mathcal{L}$. Define $T_2 = \{(c, +)\} \cup \{(x, -) : x < c\} \cup \{(c + x, +) : x \in T_1^+\} \cup \{(c + x, -) : x \in T_1^-\}$ (recall that $0 \in L$ by the definition of $\mathcal{L}$). Note that T_2 is consistent with $c + L$. Suppose that for some $c' \in \mathbb{N}_0$ and $L' \in \mathcal{L}$, $c' + L'$ is consistent with T_2. The consistency of $c' + L'$ with $\{(c, +)\} \cup \{(x, -) : x < c\}$ implies that $c' = c' + \min(L') = c$. Thus L' is consistent with $\{(x, +) : x \in T_1^+\} \cup \{(x, -) : x \in T_1^-\} = T_1$, and therefore $L' = L$. ▮ (Lemma A.1) ■

Appendix B. Proof of Proposition 4

Proposition 4. Let $\pi \in \Pi^z$ and $z \geq 2$. Then π is finitely distinguishable w.r.t. Π^z if π is equivalent to a pattern of the shape $y_1 a_1 y_2 \ldots a_k y_{k+1}$, where $a_1, \ldots, a_k \in \Sigma$ and $y_1, \ldots, y_{k+1}$ are distinct variables.

Proof We start with the case $z \geq 3$. Assume that π is of the form $y_1 a_1 y_2 \ldots a_k y_{k+1}$, where $a_1, \ldots, a_k \in \Sigma$ and $y_1, \ldots, y_{k+1}$ are distinct variables. To build a teaching set T for π w.r.t. Π^z, first put $(\pi(\varepsilon), +)$ into T. Next, for each $w \in (\mathrm{Const}(\pi(\varepsilon)))^*$ with $|w| < |\pi(\varepsilon)|$ such that $w = \pi(\varepsilon)[i_1] \ldots \pi(\varepsilon)[i_k]$ for some subsequence $(i_1, \ldots, i_k)$ of $(1, \ldots, |\pi(\varepsilon)|)$, put $(w, -)$ into T; no more than $2^{|\pi|} - 1$ of such w exist. These additional examples in T ensure that any $\pi' \in \Pi^z$

consistent with T satisfies $\pi'(\varepsilon) = \pi(\varepsilon)$. Now for each $i \in [1, k+1]$, fix some $b_i \in \Sigma$ that is different from all the constants adjacent to y_i, and put $(\beta_i, +)$ into T, where

$$\beta_i = \underbrace{a_1} \ldots \underbrace{a_{i-1}} b_i \underbrace{a_i} \cdots \underbrace{a_k} \tag{2}$$

is obtained from $\pi(\varepsilon)$ by inserting b_i between a_{i-1} and a_i. (If $i = 1$, then b_i is the first symbol of β_i; if $i = k + 1$, then b_i is the last symbol of β_i.)

Suppose π' is consistent with the examples in T so far. Suppose $A' : (X \cup \Sigma)^* \mapsto \Sigma^*$ witnesses $\beta_i \in L(\pi')$. Since $\pi'(\varepsilon) = \pi(\varepsilon)$ and $|\beta_i| = |\pi(\varepsilon)| + 1$, there is some variable y in π' that occurs exactly once in π' such that A' maps y to exactly one symbol in β_i and A' maps constants in π' to the remaining symbols in β_i. Suppose A' maps y to the symbol a_j in β_i (where the a_i's are indicated by braces in (2)) for some $j < i$. Since $\pi'(\varepsilon) = \pi(\varepsilon)$, one has that $a_{j'} = a_j$ for all $j' \in [j, i-1]$ and $a_{i-1} = b_i$. But b_i was chosen so that $b_i \neq a_{i-1}$—a contradiction. Similarly, if A' maps y to the symbol a_j in β_i (where the a_i's are indicated by braces in (2)) for some $j > i$, then one has $b_i = a_i$, which again contradicts our choice of b_i. Hence A' maps y to b_i in the decomposition (2), so that π' contains a variable y_i between a_{i-1} and a_i that occurs in π' exactly once. Repeating this argument for each $i \in [1, k+1]$ implies that π' must be of the form

$$\underbrace{X_1 y_1 X_2}\, a_1 \underbrace{X_3 y_2 X_4}\, a_2 \ldots a_k \underbrace{X_{2k+1} y_{k+1} X_{2k+2}},$$

where each y_i occurs in π' exactly once and $X_1, X_2, \ldots, X_{2k+1}, X_{2k+2} \in X^*$. But π' is equivalent to the pattern $y_1 a_1 y_2 a_2 \ldots a_k y_{k+1}$, and so $L(\pi') = L(\pi)$. Hence $\mathrm{TD}(\pi, \Pi^z) < \infty$, indeed. [3]

Now assume that $z = 2$ and let $\Sigma = \{a, b\}$. We will use the following lemma, which was shown in (Nessel and Lange, 2005, Lemma 2).

Lemma B.1 *Let $\Sigma = \{a, b\}$ and π be any pattern over $\Sigma \cup X$. Given any substring of π that has one of the following shapes: $x_i a x_j b^m x_k$, $x_i b^m x_j a x_k$, $x_i b x_j a^m x_k$ or $x_i a^m x_j b x_k$ where $m \in \mathbb{N}$, π is equivalent to the regular pattern π' obtained from π by deleting x_j.*

To keep the proof of Proposition 4 self-contained, we shall prove Lemma B.1. Suppose that $s = x_i a x_j b^m x_k$ is a substring of π; if s has one of the shapes $x_i b^m x_j a x_k$, $x_i b x_j a^m x_k$ or $x_i a^m x_j b x_k$, then a similar proof applies. Since $\pi' = \pi[x_j \to \varepsilon]$, $L(\pi') \subseteq L(\pi)$. Thus it suffices to show that for any $w \in L(\pi)$ such that w is derived from π by substituting a nonempty string for x_j, $w \in L(\pi')$. Suppose $\varphi : (\Sigma \cup X)^* \to \Sigma^*$ is a substitution witnessing $w \in L(\pi)$. We define $\phi : (\Sigma \cup X)^* \to \Sigma^*$ so that $\phi(\pi') = w$. Consider three cases.

Case (a): $\varphi(x_j) = wa^n$, where $w \in \Sigma^*$ and $n \in \mathbb{N}$. Define $\phi(x_i) = \varphi(x_i) awa^{n-1}$ and $\phi(x_l) = \varphi(x_l)$ for all $x_l \in \mathrm{Var}(\pi') \setminus \{x_i\}$.

Case (b): $\varphi(x_j) = wab^n$, where $w \in \Sigma^*$ and $n \in \mathbb{N}$. Define $\phi(x_i) = \varphi(x_i)aw$, $\phi(x_k) = b^n \varphi(x_k)$ and $\phi(x_l) = \varphi(x_l)$ for all $x_l \in \mathrm{Var}(\pi') \setminus \{x_i, x_k\}$.

Case (c): $\varphi(x_j) = b^n$, where $n \in \mathbb{N}$. Define $\phi(x_k) = b^n \varphi(x_k)$ and $\phi(x_l) = \varphi(x_l)$ for all $x_l \in \mathrm{Var}(\pi') \setminus \{x_k\}$.

3. Note that the size of the teaching set for π w.r.t. Π^z constructed in this proof is $O(2^{|\pi|})$.

▮ (Lemma B.1)

By Theorem 2, it may be assumed that π is of the form $y_1 a_1 y_2 \ldots a_k y_{k+1}$, where $a_1, \ldots, a_k \in \Sigma$ and $y_1, \ldots, y_{k+1}$ are distinct variables. To build a teaching set T for π w.r.t. Π^2, first put $(\pi(\varepsilon), +)$ into T. Next, for each $w \in (\text{Const}(\pi(\varepsilon))^*$ such that w is a proper subsequence of $\pi(\varepsilon)$, put $(w, -)$ into T; no more than $2^{|\pi|} - 1$ of such w exist. These additional examples in T ensure that any $\pi' \in \Pi^2$ consistent with T satisfies $\pi'(\varepsilon) = \pi(\varepsilon)$.

Pick $b_1, b_{k+1} \in \Sigma$ such that $b_1 \neq a_1$ and $b_{k+1} \neq a_k$. For each $i \in [2, k]$ such that $a_{i-1} = a_i$, fix $b_i \in \Sigma$ such that $b_i \neq a_i$ $(= a_{i-1})$. Define

$$\beta_i = \underbrace{a_1} \ldots \underbrace{a_{i-1}} b_i \underbrace{a_i} \ldots \underbrace{a_k} \tag{3}$$

whenever b_i is defined, and put $(\beta_i, +)$ into T. For each $i \in [2, k]$ such that $a_{i-1} \neq a_i$, put both $(a_1 \ldots a_{i-1} a a_i \ldots a_k, +)$ and $(a_1 \ldots a_{i-1} b a_i \ldots a_k, +)$ into T.

Suppose π' is consistent with the labelled examples in T so far. One can argue as in the proof for the case $z \geq 3$ that $\pi'(\varepsilon) = \pi(\varepsilon)$ and for each i such that $i \in \{1, k\}$ or $a_{i-1} = a_i$, the consistency of π' with $(\beta_i, +)$ implies that there is a free variable of π' between a_{i-1} and a_i. Now consider any $i \in [2, k]$ such that $a_{i-1} \neq a_i$. By symmetry, it may be assumed that $a_{i-1} = a$ and $a_i = b$. Suppose $A : (X \cup \Sigma)^* \mapsto \Sigma^*$ witnesses

$$\gamma_i = \underbrace{a_1} \ldots \underbrace{a_{i-1}} a \underbrace{a_i} \ldots \underbrace{a_k} \in L(\pi'). \tag{4}$$

As was argued in the proof for the case $z \geq 3$, there is some free variable y in π' such that A maps y to exactly one symbol in γ_i. Suppose A maps y to the symbol a_j in γ_i (the specific occurrence of a_j in γ_i indicated by the sequence of braces in (4)) for some $j < i$. If $a_{i-2} = a$, then (as was argued above) π' contains a free variable between a_{i-2} and a_{i-1}. If $i = 2$, then (as argued above) π' contains a free variable just before a_{i-1}. If $a_{i-2} = b$, then an argument very similar to that in the proof for the case $z \geq 3$ shows that a free variable of π' occurs either between a_{i-2} and a_{i-1} or between a_{i-1} and a_i. Further, it may be argued as in the proof for the case $z \geq 3$ that A cannot map y to any a_j in γ_i with $j \geq i$.

Suppose $B : (X \cup \Sigma)^* \mapsto \Sigma^*$ witnesses

$$\underbrace{a_1} \ldots \underbrace{a_{i-1}} b \underbrace{a_i} \ldots \underbrace{a_k} \in L(\pi'). \tag{5}$$

One can apply an argument parallel to that in the previous paragraph to show that a free variable of π' occurs either between a_i and a_{i+1} or between a_{i-1} and a_i. Thus it holds that either a free variable of π' occurs between a_{i-1} and a_i, or there exist free variables x, y of π' such that x occurs just before a_{i-1} and y occurs just after a_i; in the latter case, an application of Lemma B.1 shows that a free variable may be inserted between a_{i-1} and a_i in π', yielding a pattern that is equivalent to π'. ■

Appendix C. Example 2

Example 2. Let $\Sigma = \{a, b, c\}$ and $\pi = x_1 x_2 x_3 a x_2 x_4^2 x_5^3 x_6 b x_7 x_6 x_8$. Then π is finitely distinguishable w.r.t. Π^3 but $L(\pi)$ cannot be generated by any regular pattern.

Proof Suppose $L(\pi)$ were equal to $L(\tau)$ for some regular pattern τ. Since $|\Sigma| \geq 3$, it follows from a result in (Jiang et al., 1995) that π and τ are *similar*, that is, the constant parts of π and τ are identical and occur in the same order in the patterns, so that (after normalisation) $\tau = x_1 a x_2 b x_3$. But $acb \in L(\tau) \setminus L(\pi)$, and so $L(\tau) \neq L(\pi)$.

Now we show that $\mathrm{TD}(\pi, \Pi^3)$ is finite. We claim that $T = \{(ab, +), (a, -), (b, -), (ac^2b, +), (ac^3b, +), (acb, -), (bca^2cb, +), (acb^2ca, +)\}$ is a teaching set for π w.r.t. Π^3. Let π' be any pattern that is consistent with T. Note that the consistency of π' with $(ab, +)$, $(a, -)$ and $(b, -)$ implies that π' is of the shape $X_1 a X_2 b X_3$, where $X_1, X_2, X_3 \in X^*$. Furthermore, π' must fulfil the following conditions:

1. π' contains a variable y_1 such that y_1 occurs in X_2 exactly twice and does not occur in any other maximal variable block of π'.

2. π' contains a variable y_2 such that y_2 occurs in X_2 exactly thrice and does not occur in any other maximal variable block of π'.

3. Every variable that X_2 contains occurs in π' at least twice.

4. There is a variable y_3 that occurs in X_1 exactly once, occurs in X_2 exactly once, does not occur in X_3, and there are variables y_5 and y_6, each of which occurs in π' exactly once, such that $X_1 = Y_1 y_5 Y_2 y_3 Y_3 y_6 Y_4$ for some $Y_1, Y_2, Y_3, Y_4 \in X^*$.

5. There is a variable y_4 that occurs in X_3 exactly once, occurs in X_2 exactly once, does not occur in X_1, and there are variables y_7 and y_8, each of which occurs in π' exactly once, such that $X_3 = Z_1 y_7 Z_2 y_4 Z_3 y_8 Z_4$, where $Z_1, Z_2, Z_3, Z_4 \in X^*$.

Note that Condition 1. is implied by the consistency of π' with $\{(acb, -), (ac^2b, +)\}$, Condition 2. by the consistency of π' with $\{(acb, -), (ac^3b, +)\}$, Condition 3. by the consistency of π' with $\{(acb, -)\}$, Condition 4. by the consistency of π' with $\{(acb, -), (bca^2cb, +)\}$ and Condition 5. by the consistency of π' with $\{(acb, -), (acb^2ca, +)\}$. We claim further that any π' satisfying the preceding set of conditions generates the same language as $\pi = x_1 x_2 x_3 a x_2 x_4^2 x_5^3 x_6 b x_7 x_6 x_8$. It will be shown that $L(\pi') \subseteq L(\pi)$; the reverse inclusion may be proved similarly.

Consider any $\beta \in L(\pi')$, and let $A : (X \cup \Sigma)^* \to \Sigma^*$ be a substitution witnessing $\beta \in L(\pi')$. Note that $aA(X_2)b$ must contain a substring of the shape ac^kb for some least $k \geq 0$. In each of the following cases, we specify a substitution $\sigma : X \to \Sigma^*$ that witnesses $\beta \in L(\pi)$.

Case 1: $k = 0$. Let $\beta = \gamma_1 ab \gamma_2$, where $\gamma_1, \gamma_2 \in \Sigma^*$. Define

$$\sigma(x_i) = \begin{cases} \gamma_1 & \text{if } i = 3; \\ \gamma_2 & \text{if } i = 8; \\ \varepsilon & \text{if } i \notin \{3, 8\}. \end{cases}$$

Case 2: $k = 1$. Since every variable of X_2 occurs in π' at least twice (Condition 3.), at least one of the following cases must hold.

Case 2.1: β is of the shape $\gamma_1 c \gamma_2 acb \gamma_3$, where $\gamma_1, \gamma_2, \gamma_3 \in \Sigma^*$. Define

$$\sigma(x_i) = \begin{cases} \gamma_1 & \text{if } i = 1; \\ c & \text{if } i = 2; \\ \gamma_2 & \text{if } i = 3; \\ \gamma_3 & \text{if } i = 7; \\ \varepsilon & \text{if } i \notin \{1, 2, 3, 7\}. \end{cases}$$

Case 2.2: β is of the shape $\gamma_1 acb \gamma_2 c \gamma_3$, where $\gamma_1, \gamma_2, \gamma_3 \in \Sigma^*$. Define

$$\sigma(x_i) = \begin{cases} \gamma_1 & \text{if } i = 3; \\ c & \text{if } i = 6; \\ \gamma_2 & \text{if } i = 7; \\ \gamma_3 & \text{if } i = 8; \\ \varepsilon & \text{if } i \notin \{3, 6, 7, 8\}. \end{cases}$$

Case 3: $k > 1$. Given any $k > 1$, there are nonnegative integers m_k and n_k such that $2m_k+3n_k = k$. Let $\beta = \gamma_1 ac^k b \gamma_2$, where $\gamma_1, \gamma_2 \in \Sigma^*$. Define

$$\sigma(x_i) = \begin{cases} \gamma_1 & \text{if } i = 3; \\ c^{m_k} & \text{if } i = 4; \\ c^{n_k} & \text{if } i = 5; \\ \gamma_2 & \text{if } i = 7; \\ \varepsilon & \text{if } i \notin \{3, 4, 5, 7\}. \end{cases}$$

This completes the case distinction, showing that $\beta \in L(\pi)$. ■

Appendix D. Example 3

Example 3. Let $\Sigma = \{a, b\}$ and $\pi = x_1 x_2 a x_2 x_3^2 x_4^3 x_5 a x_5 x_6$. Then π is finitely distinguishable w.r.t. Π^2 but $L(\pi)$ cannot be generated by any regular pattern.

Proof We have already shown that $L(\pi)$ cannot be generated by any regular pattern. It remains to show that $T = \{(aa, +), (a, -), (baa, +), (aab, +), (ab^2a, +), (ab^3a, +), (aba, -), (abab, +), (ababa, +), (baba, +)\}$ is a teaching set for π w.r.t. Π^2.

Claim 1. For all patterns π', π' is consistent with T iff $L(\pi')$ consists of all finite strings $s = b^{m_1} a^{m_2} b^{m_3} a^{m_4} b^{m_5} \ldots$ such that

1. $m_2, m_4 > 0$;
2. if $m_3 = 1$, then (b occurs at least twice in $s \vee a^2$ is a substring of s).

Proof of Claim 1. Let π' be any pattern. If $L(\pi')$ consists of all finite strings $s = b^{m_1} a^{m_2} b^{m_3} a^{m_4} b^{m_5} \ldots$ satisfying Conditions 1. and 2., then one may directly verify that $\{aa, baa, aab, ab^2a, ab^3a, abab, ababa, baba\} \subset L(\pi')$ while $L(\pi') \cap \{a, aba\} = \emptyset$. Thus π' is consistent with T. Now suppose that π' is consistent with T. Then the following hold:

(i) $\big(aa \in L(\pi') \wedge a \notin L(\pi')\big) \rightarrow \pi' = X_1 a X_2 a X_3$ for some $X_1, X_2, X_3 \in X^*$.

(ii) $baa \in L(\pi') \rightarrow X_1$ contains a free variable.

(iii) $aab \in L(\pi') \rightarrow X_3$ contains a free variable.

(iv) $\big(ab^2a \in L(\pi') \wedge aba \notin L(\pi')\big) \rightarrow \pi'$ contains a variable occurring exactly twice in X_2 and not occurring in any other maximal variable block.

(v) $\big(ab^3a \in L(\pi') \wedge aba \notin L(\pi')\big) \rightarrow \pi'$ contains a variable occurring exactly thrice in X_2 and not occurring in any other maximal variable block.

(vi) $aba \notin L(\pi') \rightarrow X_2$ does not contain any free variable.

(vii) $\big(baba \in L(\pi') \wedge aba \notin L(\pi')\big) \rightarrow \pi'$ contains a variable y occurring once in X_1, once in X_2 and not occurring in any other maximal variable block.

(viii) $\big(abab \in L(\pi') \wedge aba \notin L(\pi')\big) \rightarrow \pi'$ contains a variable y occurring once in X_2, once in X_3 and not occurring in any other maximal variable block.

(ix) $\big(ababa \in L(\pi') \wedge aba \notin L(\pi')\big) \rightarrow (\pi'$ contains a variable y occurring exactly once in X_2, exactly once in X_3 and occurring in no other maximal variable block, and a free variable occurs in X_3 after the occurrence of y in $X_3) \vee (\pi'$ contains a variable y occurring exactly once in X_1, exactly once in X_2 and not occurring in any other maximal variable block, and a free variable occurs in X_1 before the occurrence of y in X_1).

First, consider any $\alpha \in L(\pi')$. By (i), α has the shape $b^{m_1}a^{m_2}b^{m_3}a^{m_4}b^{m_5}\ldots$, where $m_2, m_4 > 0$. Furthermore, if $m_3 = 1$, then (vi) implies that (b occurs at least twice in $\alpha \vee a^2$ is a substring of α). Now suppose s is a string of the shape $b^{m_1}a^{m_2}b^{m_3}a^{m_4}b^{m_5}\ldots\delta^{m_k}$ satisfying Conditions 1. and 2, where $\delta \in \{a, b\}$ and $m_i > 0$ for all $i \in \{1, \ldots, k\} \setminus \{1, 3\}$. We show that $s \in L(\pi')$ by means of the following case distinction.

Case (a): a^2 is a substring of s. Let $s = \beta_1 a^2 \beta_2$, where $\beta_1, \beta_2 \in \Sigma^*$. By (ii) and (iii), one may substitute β_1 for the free variable occurring in X_1 and β_2 for the free variable occurring in X_3.

Case (b): a^2 is not a substring of s and $m_{2j-1} \geq 2$ for some j such that $2j - 1 \leq k$. First, suppose $m_{2j-1} \geq 2$ for some j such that $2j - 1 \notin \{1, k\}$. Then $m_{2j-2}, m_{2j} \geq 1$. Let n_1 and n_2 be nonnegative integers such that $2n_1 + 3n_2 = m_{2j-1}$. By (iv) and (v), one may substitute b^{n_1} for the variable occurring twice in X_2 (and occurring in no other maximal variable block) and b^{n_2} for the variable occurring thrice in X_2 (and occurring in no other maximal variable block). By (ii) and (iii), one may substitute $b^{m_1}\ldots a^{m_{2j-2}-1}$ for the free variable occurring in X_1 and $a^{m_{2j}-1}\ldots\delta^{m_k}$ for the free variable occurring in X_3.

Second, suppose $m_{2j-1} = 1$ for all j such that $2j - 1 \notin \{1, k\}$ and $m_1 \geq 2$. By (vii), one may substitute b for the variable occurring once in X_1, once in X_2 and occurring in no other maximal variable block. By (ii) and (iii), one may substitute b^{m_1-1} for the free variable occurring in X_1 and $b^{m_5}\ldots\delta^{m_k}$ for the free variable occurring in X_3.

Third, suppose that $m_k \geq 2$ and k is odd. By (viii), one may substitute b for the variable occurring once in X_2, once in X_3 and occurring in no other maximal variable block. By (ii) and (iii), one may substitute $a^{m_1}\ldots b^{m_{k-4}}$ for the free variable occurring in X_1 and substitute b^{m_k-1} for the free variable occurring in X_3.

Case (c): s has the shape $(ba)^i b^l$ for some $i \geq 2$ and $l \in \mathbb{N}_0$. By (vii), one may substitute b for the variable occurring once in X_1, once in X_2 and occurring in no other maximal variable block. By (iii), one may substitute $b^{m_5} \ldots \delta^{m_k}$ for the free variable occurring in X_3.

Case (d): s has the shape $(ab)^i a$ for some $i \geq 2$. By (ix), at least one of the following holds: (1) one may substitute b for the variable y occurring once in X_2, once in X_3 and occurring in no other maximal variable block, and substitute $a^{m_6} \ldots \delta^{m_k}$ for the free variable in X_3 occurring after the occurrence of y in X_3, or (2) one may substitute b for the variable y occurring once in X_1, once in X_2 and occurring in no other maximal variable block, and substitute $a^{m_2} \ldots a^{m_{k-4}}$ for the free variable in X_1 occurring before the occurrence of y in X_1.

Case (e): s has the shape $(ab)^i$ for some $i \geq 2$. By (viii), one may substitute b for the variable occurring once in X_2, once in X_3 and not occurring in any other maximal variable block. By (ii), one may substitute $a^{m_2} \ldots b^{m_{k-4}}$ for the free variable occurring in X_1.

This completes the case distinction, showing that $L(\pi')$ consists of all strings s of the shape $b^{m_1} a^{m_2} b^{m_3} a^{m_4} b^{m_5} \ldots$ satisfying Conditions 1. and 2. ▮(Claim 1)

It may be directly verified that π is consistent with T. Consequently, by Claim 1, T is indeed a teaching set for π w.r.t. Π^2. ■

Appendix E. Example 4

Example 4. Let $\Sigma = \{a, b\}$ and $\pi = x_1 a x_2^2 a x_3$. Then (a) $L(\pi)$ is regular, (b) π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$, and (c) π starts and ends with variables, but π is not finitely distinguishable w.r.t. Π^2.

Proof According to (Reidenbach and Schmid, 2014, Proposition 9), $L(\pi)$ is regular; it also follows directly from the definition of π that π satisfies conditions (b) and (c). It remains to show that $\mathrm{TD}(\pi, \Pi^2) = \infty$. Suppose otherwise, and that T were a finite teaching set for $L(\pi)$ w.r.t. Π^2. Then there is an m sufficiently large so that for all $m' \geq m$, the language generated by $\pi' = x_1 a x_4^{m'} x_2^2 a x_3$ is consistent with T. Let $m' \geq m$ be odd. One has $ab^{m'}a \in L(\pi')$ via the assignment $x_1, x_2, x_3 \to \varepsilon$ and $x_4 \to b$. However, if $ab^{m'}a \in L(\pi)$ via some $B : X \to \Sigma^*$, then $B(x_1) = B(x_3) = \varepsilon$, and so $B(x_2^2) = b^{2k} = b^{m'}$ for some $k \geq 1$, which is impossible as m' is odd. ■

Appendix F. Example 5

Example 5. Let $\Sigma = \{a, b, c\}$ and $\pi = x_1 x_2 x_3 a x_2 x_4^2 x_5 b x_6 x_5 x_7$. Then (a) $L(\pi)$ is regular, (b) π does not contain any substring $\alpha \in \Sigma^+$ such that $|\alpha| \geq 2$, and (c) π starts and ends with variables, but π is not finitely distinguishable w.r.t. Π^3.

Proof According to (Jain et al., 2010, Theorem 2), $L(\pi)$ is regular; also, by definition, π satisfies (b) and (c). Now assume that T were a finite teaching set for $L(\pi)$ w.r.t. Π^3. As in Example 4, there is an m large enough so that whenever $m' \geq m$, $\pi' = x_1 x_2 x_3 a x_8^{m'} x_2 x_4^2 x_5 b x_6 x_5 x_7$ is consistent with T. Fix some odd $m' \geq m$. Note that the assignment $x_1, x_2, x_3, x_4, x_5, x_6, x_7 \to \varepsilon, x_8 \to c$ witnesses $ac^{m'}b \in L(\pi')$. If, however, there were some assignment $B : X \to \Sigma^*$ witnessing $ac^{m'}b$, then it must hold that $B(x_1) = B(x_2) = B(x_3) = B(x_5) = B(x_6) = B(x_7) = \varepsilon$ and $B(x_4^2) = c^{2k} = c^{m'}$ for some $k \geq 1$, contradicting the fact that m' is odd. ■

Appendix G. Proof of Theorem 6

Theorem 6. Let $1 \leq z < \infty$ and $\pi \in \Pi^z$. If π is finitely distinguishable w.r.t. Π^z, then $L(\pi)$ is regular.

Proof Let $\Sigma = \{a_1, \ldots, a_z\}$. For each $\delta \in \Sigma$ and $w \in (X \cup \Sigma)^*$, let $\#(\delta)[w]$ denote the number of occurrences of δ in w. Further, for any $\beta, \gamma \in \Sigma^*$, recall that the *shuffle product* of β and γ, denoted by $\beta \sqcup\!\sqcup \gamma$, is the set $\{\beta_1\gamma_1\beta_2\gamma_2 \ldots \beta_k\gamma_k : \beta_i, \gamma_i \in \Sigma^* \wedge \beta_1\beta_2 \ldots \beta_k = \beta \wedge \gamma_1\gamma_2 \ldots \gamma_k = \gamma\}$, and the shuffle product of two sets S and T, denoted by $S \sqcup\!\sqcup T$, is the set $\bigcup_{s \in S \wedge t \in T} s \sqcup\!\sqcup t$ (Lothaire, 1983).

Suppose T were a finite teaching set for π w.r.t. Π^z. Fix some $m > \max\{|\alpha| : \alpha \in T^+ \cup T^- \vee |\alpha| = |\pi|\}$. Consider any pair $(I, J) \in \wp([1, z]) \times \wp([1, z])$ such that $I \cap J = \emptyset$. Let $I = \{i_1, \ldots, i_k\}$ and $J = \{j_1, \ldots, j_\ell\}$. Define

$$
\begin{aligned}
S_I = \{w \in L(\pi) : (\forall 1 \leq d \leq k)[\#(a_{i_d})[\pi] + 1 \leq \#(a_{i_d})[w] \leq \#(a_{i_d})[\pi] + m \\
- 1] \wedge (\forall e \in [1, z] \setminus I)[\#(a_e)[w] = \#(a_e)[\pi]]\}, \\
T_J = \{v \in \{a_{j_1}, \ldots, a_{j_\ell}\}^* : (\forall 1 \leq d \leq \ell)[\#(a_{j_d})[v] = m]\}.
\end{aligned}
$$

Given S_I and T_J, set $E_{I,J} = (S_I \sqcup\!\sqcup T_J) \sqcup\!\sqcup \{a_{j_1}, \ldots, a_{j_\ell}\}^*$. Observe that S_I and T_J are both finite and hence regular, while $\{a_{j_1}, \ldots, a_{j_\ell}\}^*$ is also regular. As the shuffle operation preserves regularity, it follows that $E_{I,J}$ is regular. Further, since the regular languages are closed under the union operation, the required result follows immediately from the next claim.

Claim 1. $L(\pi) = \bigcup_{I,J \subseteq [1,z] \wedge I \cap J = \emptyset} E_{I,J}$.

Proof of Claim 1. We first show that $L(\pi) \subseteq \bigcup_{I,J \subseteq [1,z] \wedge I \cap J = \emptyset} E_{I,J}$. Consider any $\alpha \in L(\pi)$. Define $I = \{d : \#(a_d)[\pi] + 1 \leq \#(a_d)[\alpha] \leq \#(a_d)[\pi] + m - 1\}$ and $J = \{e : \#(a_e)[\alpha] \geq \#(a_e)[\pi] + m\}$. Then $\alpha \in E_{I,J}$.

Now it is shown that $\bigcup_{I,J \subseteq [1,z] \wedge I \cap J = \emptyset} E_{I,J} \subseteq L(\pi)$. Choose any $I, J \subseteq [1, z]$ such that $I \cap J = \emptyset$. Let $I = \{i_1, \ldots, i_k\}$ and $J = \{j_1, \ldots, j_\ell\}$. Pick any $\alpha \in S_I, \beta \in T_J$ and $\gamma \in \{a_{j_1}, \ldots, a_{j_\ell}\}^*$. One has to show that for any $w \in (\alpha \sqcup\!\sqcup \beta) \sqcup\!\sqcup \gamma$, $w \in L(\pi)$. Let $\varphi : X \mapsto \Sigma^*$ be a substitution witnessing $\alpha \in L(\pi)$. Since $w \in (\alpha \sqcup\!\sqcup \beta) \sqcup\!\sqcup \gamma$, there is some $v \in \{a_{j_1}, \ldots, a_{j_\ell}\}^*$ such that whenever $1 \leq d \leq \ell$, a_{j_d} occurs at least m times in v and

$$
w = v_1\alpha_1 v_2\alpha_2 \ldots v_{n-1}\alpha_{n-1}v_n \tag{6}
$$

for some $v_1, \ldots, v_n \in \{a_{j_1}, \ldots, a_{j_\ell}\}^*$ and $\alpha_1, \ldots, \alpha_{n-1} \in \Sigma^*$ with $\alpha = \alpha_1\alpha_2 \ldots \alpha_{n-1}$ and $v = v_1 \ldots v_n$.

One now derives a pattern τ from the decomposition (6) of w as follows. Let $\pi_1, \ldots, \pi_{n-1} \in (X \cup \Sigma)^*$ be strings such that $\pi = \pi_1 \ldots \pi_{n-1}$ and $\varphi(\pi_i) = \alpha_i$ for all $i \in [1, n-1]$. Replace each α_i (here we are referring to the specific occurrence of α_i starting at the $(|v_1\alpha_1 \ldots v_i| + 1)^{st}$ position of w) with π_i. Next, choose distinct variables $y_1, \ldots, y_\ell \notin \text{Var}(\pi)$. For each $d \in [1, \ell]$, substitute y_d for every occurrence of a_{j_d} in $v_1, v_2, \ldots, v_n$ (as before, for every $i \in [1, n]$, we are referring to the specific occurrence of v_i starting at the $(|v_1 \ldots \alpha_{i-1}| + 1)^{st}$ position of w). Note that τ can be derived from π by interleaving π with a string consisting of the variables $y_1, \ldots, y_\ell$, and therefore

$L(\pi) \subseteq L(\tau)$. Further, τ is consistent with T because every additional variable y_i occurs at least m times in τ. Thus $L(\tau) \subseteq L(\pi)$, and as $w \in L(\tau)$, it follows that $w \in L(\pi)$. ■

Appendix H. Proof of Theorem 7

Theorem 7. Let $z \in \{2, 3\}$, $\Sigma_1 = \{a, b\}$, $\Sigma_2 = \{a, b, c\}$ and $\pi = X_1 c_1 X_2 c_2 \ldots X_{n-1} c_{n-1} X_n$, where $X_2, \ldots, X_{n-1} \in X^+$, $c_1, \ldots, c_{n-1} \in \Sigma_1^+$ if $z = 2$, $c_1, \ldots, c_{n-1} \in \Sigma_2^+$ if $z = 3$, and $X_1, X_n \in X^*$. If π is finitely distinguishable w.r.t. Π^z, then the following conditions hold for all $i \in [1, n-1]$.

1. If $z = 2$, then $c_i \in \{a, b, ab, ba\}$; if $z = 3$, then $c_i \in \Sigma_2$.
2. If $z = 2$, then for all $\alpha \in \{X_1, X_n, \delta X_i \delta, \delta X_i \overline{\delta} X_{i+1} \delta, \delta\overline{\delta} X_i \delta, \delta X_i \overline{\delta}\delta\}$ such that α is a substring of π, where $\delta, \overline{\delta} \in \Sigma$ and $\delta \neq \overline{\delta}$, there is a $k \geq 1$ for which α contains variables $y_1, \ldots, y_k$ such that for all $j \in [1, k]$, y_j occurs q_j times in α for some $q_j \geq 1$, y_j does not occur outside the block α and $\gcd(q_1, \ldots, q_k) = 1$. If $z = 3$, then the latter statement holds for $\alpha = X_i$.
3. If $z = 2$, then π contains at least one free variable; if $z = 3$, then X_1 and X_n each contains at least one free variable.

Proof Let T be a finite teaching set for $L(\pi)$ w.r.t. Π^z and fix any $m > \max(\{|\gamma| : \gamma \in T^+ \cup T^-\} \cup \{|\pi|\})$.

Proof of (1). Let $z = 2$. Suppose $\pi[i]\pi[i+1] = aa$ for some $i \in [1, |\pi| - 1]$. Choose some variable $y \notin \text{Var}(\pi)$, and let π' be the pattern obtained from π by inserting y^m between the i^{th} and $(i+1)^{st}$ positions of π. Note that π' is consistent with T. Furthermore, let β be the string derived from π' by substituting b for y and ε for every other variable. Since the number of times that aa occurs in β is strictly less than the number of times it occurs in π, one has $\beta \in L(\pi') \setminus L(\pi)$, a contradiction.

Now suppose $\pi[i]\pi[i+1]\pi[i+2] = aba$ for some $i \in [1, |\pi| - 2]$. Let π'' be the pattern obtained from π by inserting y^m between the i^{th} and $(i+1)^{st}$ positions of π, and let θ be the string derived from π'' by substituting b for y and ε for every other variable. One may verify as in the earlier case that π'' is consistent with T but $\theta \in L(\pi'') \setminus L(\pi)$.

If $z = 3$, then the proof that $c_i \in \Sigma_2$ is similar to the preceding proof.

Proof of (2). Let $z = 2$. First consider the case $\alpha = X_1$. Choose $\delta \in \Sigma$ so that δ is different from the first symbol of c_1. As before, choose a variable $y \notin \text{Var}(\pi)$, and note that for all $j \geq m$, $y^j \pi$ is consistent with T. Thus $\delta^j \pi(\varepsilon) \in L(\pi)$ for all $j \geq m$. This implies that $X_1 \neq \varepsilon$, and that there exist variables $y_1, \ldots, y_k$ occurring only in X_1 such that for all $j \geq m$, there are nonnegative integers $m_1, \ldots, m_k$ for which $\sum_{i=1}^{k} m_i q_i = j$, where q_i is the number of times that y_i occurs in X_1. Therefore $\gcd(q_1, \ldots, q_k) = 1$. The case $\alpha = X_n$ can be handled similarly.

Now suppose $\alpha = aX_i a = \pi[j]\pi[j+1] \ldots \pi[j+l]$. Choose some variable $y \notin \text{Var}(\pi)$, and for any $m' \geq m$ let $\pi_{m'}$ be the pattern obtained from π by inserting $y^{m'}$ between the j^{th} and $(j+1)^{st}$ positions of π. Let $\beta_{m'}$ be the string derived from $\pi_{m'}$ by substituting b for y and ε for all other variables. As in the previous case, note that $\pi_{m'}$ is consistent with T and so $\beta_{m'} \in L(\pi)$, which means that there exist variables $y_1, \ldots, y_k$ occurring only in X_i such that if q_i is the number of times that y_i occurs in X_i, then $\gcd(q_1, \ldots, q_k) = 1$.

Finally, let $\alpha = aX_ibX_{i+1}a = \pi[j_1]\ldots\pi[j_1 + l_1]$. The proof is very similar to that of the previous case; here one defines for every $m' \geq m$ the pattern $\pi_{m'}$ obtained from π by inserting $y^{m'}$ between the j_1^{th} and $(j_1+1)^{st}$ positions of π and setting $\beta_{m'}$ to be the string derived from $\pi_{m'}$ by replacing y with b and every other variable with ε. The remaining cases in (2) (including the case $z = 3$) can be dealt with similarly.

Proof of (3). Let $z = 2$. Choose two distinct variables $y_1, y_2 \notin \text{Var}(\pi)$, and define

$$\tau = \pi\, \underbrace{y_1^m y_2^m y_1^m}\, \underbrace{y_1^{m+1} y_2^{m+1} y_1^{m+1}} \ldots \underbrace{y_1^{4m} y_2^{4m} y_1^{4m}}. \tag{7}$$

Let β be the string derived from τ by substituting a for y_1, b for y_2, and ε for all other variables; that is,

$$\beta = \pi(\varepsilon)\, \underbrace{a^m b^m a^m}\, \underbrace{a^{m+1} b^{m+1} a^{m+1}} \ldots \underbrace{a^{4m} b^{4m} a^{4m}}. \tag{8}$$

Since τ is consistent with T, one has that $\beta \in L(\pi)$. Let $A : (X \cup \Sigma)^* \to \Sigma^*$ be a substitution witnessing $\beta \in L(\pi)$. By statement (1), each constant block of π overlaps with at most one substring of the form $a^{m+i}b^{m+i}a^{m+i}$. Further, there is some $j \in [0, 3m]$ such that for some $z \in \text{Var}(\pi)$, A maps an occurrence of z in π to a substring β' of β such that $a^{m+j}b^{m+j}a^{m+j}$ (whose specific occurrence in β is indicated by braces in (8)) is a substring of β'; otherwise, for each occurrence of a variable z' in π', A maps this occurrence of z' to a substring of $a^{m+i}b^{m+i}a^{m+i}a^{m+i+1}b^{m+i+1}a^{m+i+1}$ (whose specific occurrence in β is indicated by braces in (8)) for at most one $i \in [0, 3m-1]$, and so $|A(\pi)| < \beta$, a contradiction. Since β cannot contain two copies of $a^{m+j}b^{m+j}a^{m+j}$, z must be a free variable of π, as required. The fact that X_1 and X_n each contains at least one free variable if $z = 3$ can be proven similarly. ∎

Appendix I. Proof of Theorem 8

Theorem 8. Let $z \in \mathbb{N} \cup \{\infty\}$ and let π be a regular pattern over Σ. Then $\text{TD}(\pi, \text{R}\Pi^z) \leq 2|\pi| + 1$.

Proof It will be shown later (Theorem 12) for all regular patterns π, $\text{TD}(\pi, \text{R}\Pi^z) \leq 3$ when $z = 1$ and $\text{TD}(\pi, \text{R}\Pi^z) \leq 5$ when $z \geq 7$. We shall therefore assume that $2 \leq z \leq 6$. A teaching set T for π w.r.t. $\text{R}\Pi^z$ may be constructed as follows. Let $w = \pi(\varepsilon)$. First, put $(w, +)$ into T. Second, for each $i \in [1, |w|]$, fix some $a_i \in \Sigma$ such that $a_i \neq w[i]$ (which is possible because $z \geq 2$), let w_i be the string derived from w by replacing $w[i]$ with a_i, and put $(w_i, -)$ into T. Let τ be any regular pattern that is consistent with the labelled examples put into T so far, and observe that $\tau(\varepsilon) = w$. Without loss of generality, one may assume that τ has the shape $c_1x_1c_2\ldots c_n$, where $c_1, c_n \in \Sigma^*$ and $c_2, \ldots, c_{n-1} \in \Sigma^+$. To finish the construction of T, the cases (i) $z = 2$ and (ii) $3 \leq z \leq 6$ will be considered separately.

Case (i): $z = 2$. Let $\Sigma = \{a, b\}$. We will apply Lemma B.1 several times in this proof.

Define $(p_1, p_2, \ldots, p_{|w|})$ to be the sequence of position numbers of π such that for all $i \in \{1, \ldots, |w|\}$, $\pi[p_i] = w[i]$. Similarly, define $(q_1, q_2, \ldots, q_{|w|})$ to be the sequence of position numbers of τ such that for all $i \in \{1, \ldots, |w|\}$, $\tau[q_i] = w[i]$. Note that since π and τ are assumed to have the shape $c_1x_1c_2x_2\ldots x_{n-1}c_n$, where $c_1, c_2 \in \Sigma^*$ and $c_2, \ldots, c_{n-1} \in \Sigma^+$, it holds that for all $i \in \{1, \ldots, |w|\}$, either $p_{i+1} = p_i + 1$ (resp. $q_{i+1} = q_i + 1$) (no variable of

π (resp. τ) occurs between $w[i]$ and $w[i+1]$) or $p_{i+1} = p_i + 2$ (resp. $q_{i+1} = q_i + 2$) (exactly one variable of π (resp. τ) occurs between $w[i]$ and $w[i+1]$). By applying Lemma B.1 as often as necessary, one may assume that π and τ possess the following property.

Property 1. Suppose that for some $\alpha \in (\Sigma \cup X)^*$, $m \geq 1$ and distinct variables x_i and x_j, $x_i a^m \alpha b x_j$ is a substring of π (resp. τ). If b does not occur in α, then α contains at least one variable. A similar statement holds with any of the strings in $\{x_i b^m \alpha x_j, x_i a \alpha b^m x_j, x_i a \alpha b^m x_j, x_i b \alpha a^m x_j\}$ substituted for $x_i a^m \alpha b x_j$.

In other words, if π (resp. τ) contains a substring of the shape $x_i a^m b x_j$, where $m \geq 1$ and x_i and x_j are distinct variables, then one can extend π (resp. τ) by inserting a new variable between a^m and b. Note that one can only add a finite number of new variables to π since it is assumed throughout this proof that the regular patterns are always expressed as $c_1 x_1 c_2 x_2 \dots x_{n-1} c_n$, where $c_1, c_n \in \Sigma^*$ and $c_2, \dots, c_{n-1} \in \Sigma^+$. The remaining elements of T are defined as follows.

1. Add two labelled examples that identify the starting and ending symbols of π. Fix some $v_1 \in \Sigma \setminus \{w[1]\}$. If $p_1 = 1$, that is, π starts with a constant, then put $(v_1 w, -)$ into T. If $p_1 = 2$, that is, π starts with a variable, then put $(v_1 w, +)$ into T. If τ were consistent with T, then τ starts with a variable iff π starts with a variable. Similarly, fix some $v_2 \in \Sigma \setminus \{w[|w|]\}$; if $p_{|w|} = |\pi|$, then put $(wv_2, -)$ into T, and if $p_{|w|} = |\pi| - 1$, then put $(wv_2, +)$ into T. If τ were consistent with T, then τ ends with a variable iff π ends with a variable.
2. Now consider any substring $w[i]w[i+1]$ of w such that $w[i] = w[i+1]$. Fix some $a_i \in \Sigma \setminus \{w[i]\} = \Sigma \setminus \{w[i+1]\}$. Let w' be the string obtained from w by inserting a_i between $w[i]$ and $w[i+1]$. If $p_{i+1} = p_i + 2$, then put $(w', +)$ into T; if $p_{i+1} = p_i + 1$, then put $(w', -)$ into T. Suppose that $(w', +) \in T$. We argue that if τ were consistent with T, then $q_{i+1} = q_i + 2$. Since $\tau(\varepsilon) = w$ and $|w'| = |w| + 1$, w' is derived from τ by replacing exactly one variable x_j with a constant symbol. Let $\varphi : (\Sigma \cup X)^* \to \Sigma^*$ be a substitution witnessing $w' \in L(\tau)$. Suppose φ maps x_j to the $(j')^{th}$ position of w' for some $j' \leq i$. Since $\tau(\varepsilon) = \pi(\varepsilon) = w$, it follows that $w'[l+1] = w[l]$ for all $l \geq j'$, contradicting the fact that $w'[i+1] \neq w[i]$. If φ maps x_j to the $(j'')^{th}$ position of w' for some $j'' \geq i+2$, then $w'[i+1] = w[i]$, which again yields a contradiction. Hence x_j occurs between q_i and q_{i+1}, that is, $q_{i+1} = q_i + 2$. One can argue similarly that if $(w', -) \in T$ and τ were consistent with T, then $q_{i+1} = q_i + 1$.
3. Next, add a labelled example to T so that a variable of π occurs between $w[1]$ and $w[2]$ iff a variable of τ occurs between $w[1]$ and $w[2]$. Suppose that $p_2 = p_1 + 2$, that is, a variable of π occurs between $w[1]$ and $w[2]$. The case $w[1] = w[2]$ was handled in 2. By symmetry of a and b, it may be assumed that $w[1] = a$ and $w[i] = b$ for all $2 \leq i \leq m$, where either $m = |w|$ or $w[m+1] = a$. If π and τ do not start with variables, then let u_1 be the string obtained from w by inserting a between $w[1]$ and $w[2]$, and put $(u_1, +)$ into T. The consistency of τ with T would imply that $q_2 = q_1 + 2$. Suppose π and τ both start with variables. In Step 2., we added an example to T so that for any $j, j+1$ with $2 \leq j, j+1 \leq m$, a variable of π occurs between $w[j]$ and $w[j+1]$ iff a variable of τ occurs between $w[j]$ and $w[j+1]$. If a variable of π (resp. τ) occurs between $w[j]$

and $w[j+1]$ for some j such that $2 \le j, j+1 \le m$, then by Lemma B.1 a variable of π (resp. τ) occurs between $w[1]$ and $w[2]$. If no variable of π (resp. τ) occurs between $w[j]$ and $w[j+1]$ whenever $2 \le j, j+1 \le m$, then let u_2 be the string obtained from w by inserting b between $w[1]$ and $w[2]$, and put $(u_2, +)$ into T. The consistency of τ with T then implies that a variable of τ occurs either between $w[1]$ and $w[2]$ or just after $w[m]$; note that the latter case also implies that a variable of τ occurs between $w[1]$ and $w[2]$. An analogous argument holds if $p_2 = p_1 + 1$. Similarly, add a labelled example to T so that a variable of π occurs between $w[|w|-1]$ and $w[|w|]$ iff a variable of τ occurs between $w[|w|-1]$ and $w[|w|]$.

4. Finally, consider any substring of w of the shape $s = w[i]w[i+1]w[i+2]w[i+3]$. We would like to add a labelled example to T so that $p_{i+2} = p_{i+1} + 2$ iff $q_{i+2} = q_{i+1} + 2$ (that is, a variable of π occurs between $w[i+1]$ and $w[i+2]$ iff a variable of τ occurs between $w[i+1]$ and $w[i+2]$). The case $w[i+1] = w[i+2]$ was handled in Step 2. By symmetry of a and b, it may be assumed that one of Subcases (1)–(4) holds; in each subcase, suppose that $p_{i+2} = p_{i+1} + 2$.

Subcase (1): $s = abaa$. Let t_1 be the string obtained from w by inserting ba between $w[i+1]$ and $w[i+2]$, and put $(t_1, +)$ into T.

Claim 1. If τ were consistent with T, then at least one of the following would hold: $q_{i+2} = q_{i+1} + 2$, or variables of τ occur between $w[i]$ and $w[i+1]$ as well as between $w[j]$ and $w[j+1]$ for some $j \ge i+2$ such that $w[j'] = a$ for all $j' \in [i+2, j]$.

Proof of Claim 1. Let $\phi : (\Sigma \cup X)^* \to \Sigma^*$ be a substitution witnessing $t_1 \in L(\tau)$. Since $|t_1| = |w| + 2$, and $\tau(\varepsilon) = w$, one of the following cases holds.

Case (a): There is exactly one variable x_k of τ such that for some $j \in [1, |t_1| - 1]$, ϕ maps x_k to $t_1[j]t_1[j+1]$. If $j \le i$, then $t_1[j]t_1[j+1] = t_1[j+2l]t_1[j+2l+1]$ for all l such that $j+2 \le j+2l, j+2l+1 \le i+4$, which is impossible since $t_1[i]t_1[i+1]t_1[i+2]t_1[i+3] = abba$. If $j = i+1$, then $a = t_1[i+3] = w[i+1] = b$, a contradiction. Similarly, if $j \ge i+3$, then $b = t_1[i+2] = w[i+2] = a$, a contradiction. Hence $j = i+2$.

Case (b): There are distinct variables x_k, x_l such that ϕ maps x_k to $t_1[j_1]$ and ϕ maps x_l to $t_1[j_2]$ for some $j_1, j_2 \in [1, |t_1|]$ such that $j_2 > j_1 + 1$. Suppose $j_1 < i+2$. First, suppose that $t_1[j_1] = a$. Then either $t_1[j'] = a$ for all $j' \in [j_1, i+1]$ (which is impossible) or $j_2 \in [j_1+2, i+1]$, $t_1[j_2] = b$ and $t_1[j_2+2h-1]t_1[j_2+2h] = ab$ for all $h \ge 1$ such that $j_2+1 \le j_2+2h-1, j_2+2h \le i+3$, which is impossible because $t_1[i]t_1[i+1]t_1[i+2]t_1[i+3] = abba$. Second, suppose that $t_1[j_1] = b$. If $j_1 \le i$, then either $t_1[j'] = b$ for all $j' \in [j_1, i+1]$ or $j_2 \in [j_1+1, i+1]$ and $t_1[j_2+2h-1]t_1[j_2+2h] = ba$ for all $h \ge 1$ such that $j_2+1 \le j_2+2h-1, j_2+2h \le i+3$, a contradiction. Furthermore, if $j_1 \ge i+3$, then $b = t_1[i+2] = w[i+2] = a$, a contradiction. Consequently, $j_1 \in \{i+1, i+2\}$.
Now suppose $j_2 \ge i+6$. Suppose that $t_1[j_2] = b$. Then for all $j_3 \in [i+4, j_2-1]$, $t_1[j_3] = b$, which is impossible since $t_1[i+4]t_1[i+5] = aa$. Hence we may assume that $t_1[j_2] = a$. Then for all $j_3 \in [i+6, j_2-1]$, $t_1[j_3] = a$.
It follows that either x_k occurs between $w[i+1]$ and $w[i+2]$, that is, $q_{i+2} = q_{i+1} + 2$, or x_k occurs between $w[i]$ and $w[i+1]$ and x_l occurs between $w[j]$

and $w[j+1]$ for some $j \geq i+2$ such that $w[j'] = a$ for all $j' \in [i+2, j]$. ∎ (Claim 1)

Note that if variables of τ occur between $w[i]$ and $w[i+1]$ as well as between $w[j]$ and $w[j+1]$ for some $j \geq i+2$ such that $w[j'] = a$ for all $j' \in [i+2, j]$, then Lemma B.1 implies that a variable of τ must occur between $w[i+1]$ and $w[i+2]$.

Subcase (2): $s = bbab$. Let t_2 be the string obtained from w by inserting ba between $w[i+1]$ and $w[i+2]$, and put $(t_2, +)$ into T. One can argue similarly to Subcase (1) that a variable of τ must occur between $w[i+1]$ and $w[i+2]$.

Subcase (3): $s = bbaa$. Let t_3 be the string obtained from w by inserting ab between $w[i+1]$ and $w[i+2]$, and put $(t_3, +)$ into T. The rest of the argument proceeds analogously to Subcase (1).

Subcase (4): $s = abab$. Let t_4 be the string obtained from w by inserting ba between $w[i+1]$ and $w[i+2]$, and put $(t_4, +)$ into T. The rest of the argument proceeds analogously to Subcase (1).

The case $p_{i+2} = p_{i+1} + 1$ can be handled analogously to Subcases (1)–(4).

T now contains a total of $2|\pi| + 1$ labelled examples, and this completes the proof of Case (i).

Case (ii): $3 \leq z \leq 6$. For each pair of adjacent constants $w[i], w[i+1]$ such that $1 \leq i, i+1 \leq |w|$, fix some $a_i \in \Sigma \setminus \{w[i], w[i+1]\}$ (which is possible because $|\Sigma| \geq 3$) and let s_i be the string derived from w by inserting a_i between $w[i]$ and $w[i+1]$. Put $(s_i, +)$ into T if $p_{i+1} = p_i + 2$ and put $(s_i, -)$ into T if $p_{i+1} = p_i + 1$. Fix some $b_1 \in \Sigma \setminus \{w[1], w[|w|]\}$. Set $\alpha = b_1 w$ and $\beta = w b_1$. Put $(\alpha, +)$ into T if π starts with a variable and put $(\alpha, -)$ into T if π starts with a constant. Put $(\beta, +)$ into T if π ends with a variable and put $(\beta, -)$ into T if π ends with a constant. One can argue similarly to Step 2 in the proof of Case (i) that if τ were consistent with T, then $L(\tau) = L(\pi)$.

∎

Appendix J. Proof of Theorem 9

Theorem 9. Let $z \in \mathbb{N} \cup \{\infty\}$ and let π be a 1-variable pattern over Σ. Then $\mathrm{TD}(\pi, 1\Pi^z) < \infty$ iff π contains a variable. If π contains a variable, then $\mathrm{TD}(\pi, 1\Pi^z) = O(|\pi|)$ if $z = 1$ and $\mathrm{TD}(\pi, 1\Pi^z) = O(|\pi|^3)$ if $z \geq 2$ (including $z = \infty$).

Proof We prove the second part of the statement. Suppose that π contains a variable.

Case (i): $z = 1$. Let $\Sigma = \{a\}$ and $\pi = a^m x^n$. A teaching set for π w.r.t. $1\Pi^1$ is $\{(a^x, -) : x < m\} \cup \{(a^m, +), (a^{m+n}, +)\} \cup \{(a^{m+x}, -) : 0 < x < n\}$. Note that $\{(a^m, +)\} \cup \{(a^x, -) : x < m\}$ uniquely identifies a^m as the constant part of π, while $\{(a^{m+n}, +)\} \cup \{(a^{m+x}, -) : 0 < x < n\}$ uniquely identifies the variable block of π among all π' such that $\pi'(\varepsilon) = a^m$.

Case (ii): $z \geq 2$ (including $z = \infty$). Let $\pi = c_1 X_1 c_2 X_2 \ldots X_{n-1} c_n$, where $c_1, c_n \in \Sigma^*$, $c_2, \ldots, c_{n-1} \in \Sigma^+$ and $X_1, \ldots, X_{n-1} \in \{x\}^+$. Build a teaching set T as follows. First, choose any two distinct $a, b \in \Sigma$. Put $(\pi(a), +)$ and $(\pi(b), +)$ into T. Let π' be any 1-variable pattern that is consistent with $\{(\pi(a), +), (\pi(b), +)\}$. Note that since $|\pi(a)| = |\pi(b)|$

and π' contains at most one variable (with possibly more than one occurrence), any substitutions $\varphi_1, \varphi_2 : (\Sigma \cup X)^* \to \Sigma^*$ such that $\varphi_1(\pi') = \pi(a)$ and $\varphi_2(\pi') = \pi(b)$ satisfy $\varphi_1^{-1}(\pi(a)[i]) = \varphi_2^{-1}(\pi(b)[i])$ for all $i \in [1, |\pi(a)|]$. In particular, consider any $j \in [1, |\pi|]$ such that $\pi[j]$ is a variable; since $\pi(a)[j] = a \neq b = \pi(b)[j]$, $\varphi_1^{-1}(\pi(a)[j])$ is also a variable. Further, let $\pi' = d_1 Y_1 d_2 Y_2 \ldots Y_{k-1} d_k$, where $d_1, d_k \in \Sigma^*$, $d_2, \ldots, d_{k-1} \in \Sigma^+$ and $Y_1, \ldots, Y_{k-1} \in \{x\}^+$. Consider the following decomposition of $\pi(a)$:

$$\underbrace{c_1}\, a^{|X_1|} \underbrace{c_2}\, a^{|X_2|} \ldots a^{|X_{n-1}|} \underbrace{c_n} \tag{9}$$

There is a sequence $(i_1, \ldots, i_k)$ such that $1 \leq i_1 \leq \ldots \leq i_k \leq n$ and φ_1 maps d_j to c_{i_j} for all $j \in [1, k]$ (where the c_i's are indicated by braces in the decomposition (9)). Further, for every $j \in [1, k]$, $i_j < i_{j+1}$. To see this, assume to the contrary that there exists some $l \in [1, k]$ such that $i_l = i_{l+1} = m$ for some $m \in [1, n]$. Then φ_1 and φ_2 both map Y_{i_l} to the same proper substring of c_m. As $Y_{i_l} = x^u$ for some $u \geq 1$, it follows that $\varphi_1(x) = \varphi_2(x)$ and therefore $\varphi_1(\pi') = \varphi_2(\pi')$, a contradiction. Thus $i_1 < \ldots < i_k$ indeed holds. Further, for every $i \in [1, n]$, there are $O(|\pi|^2)$ substrings of c_i. Consequently, since $n \leq |\pi|$, $|\{\tau(\varepsilon) : \tau \in (\Sigma \cup X)^+ \wedge \tau \text{ is consistent with } T\}| = O(|\pi|^3)$. For each $w \in \{\tau(\varepsilon) : \tau \in (\Sigma \cup X)^+ \wedge \tau \text{ is consistent with } T\}$ such that $w \neq \pi(\varepsilon)$, put $(w, -)$ into T. Hence if π' is consistent with T, then $\pi'(\varepsilon) = \pi(\varepsilon)$. In addition, π' has the shape $c_1 Y_1 \ldots Y_{n-1} c_n$, where $Y_1, \ldots, Y_{n-1} \in \{x\}^+$ and for some $\mu \geq 1$, $|X_i| = \mu |Y_i|$ for all $i \in [1, n-1]$. Fix some $a \in \Sigma$, and for each possible choice of $\mu > 1$, put the negative example $\left(c_1 a^{\frac{|X_1|}{\mu}} \ldots a^{\frac{|X_{n-1}|}{\mu}} c_n, -\right)$ into T. There are at most $|\pi|$ possible choices of $\mu > 1$. At this stage, T contains $2 + O(|\pi|^3) + |\pi| = O(|\pi|^3)$ examples and every $\pi' \in 1\Pi^z$ consistent with T must be equivalent to π.

■

Appendix K. Proof of Theorem 10(2)

Theorem 10(2). Let $z \in \mathbb{N} \cup \{\infty\}$ and let $\pi = x_1^{n_1} \ldots x_k^{n_k}$ be a non-cross pattern over Σ. If $z \geq 2$, then $\mathrm{TD}(\pi, \mathrm{NC}\Pi^z) < \infty$ iff $n_i = 1$ for some $i \in [1, k]$, i.e., iff π contains at least one non-repeated variable.

Proof Define n_{k+1} and π' as in the earlier proof sketch of Theorem 10(2). Again, let $m_1, \ldots, m_k$, m_{k+1} be a sequence such that $m_i n_i < m_{i+1} n_{i+1}$ for all $i \leq k$. We show that the string w obtained from π' by replacing every odd-indexed variable x_{2i-1} with $a^{m_{2i-1}}$ and every even-indexed variable x_{2i} with $b^{m_{2i}}$ is in $L(\pi') \setminus L(\pi)$. That $w \in L(\pi')$ follows directly by construction; we focus on proving $w \notin L(\pi)$. Suppose for a contradiction that some substitution $\varphi : X \to \Sigma^*$ witnesses $w \in L(\pi)$. As in the proof of Theorem 3(3), Case (i.1), the morphism extending φ induces a mapping I_φ from the set of all intervals of positions of π to the set of all intervals of positions of w. For any $i, j \in \{1, \ldots, |w|\}$ with $i \leq j$, let $w[i : j]$ denote the specific factor of w from its i^{th} position to its j^{th} position. For all $j, \ell \in \{1, \ldots, |w|\}$ with $j \leq \ell$ and $i \in \{1, \ldots, k\}$, say that $w[j : \ell]$ *cuts* $\varphi(x_i^{n_i})$ iff I_φ maps the interval of positions of π corresponding to the (unique) occurrence of $x_i^{n_i}$ in π to a nonempty interval $[i', j']$ such that one of the following holds: (1) $i' < i$ and $j' \geq i$, or (2) $i' \leq j$

and $j' > j$. In other words, $w[j : \ell]$ cuts $\varphi(x_i^{n_i})$ iff I_φ maps the interval corresponding to $x_i^{n_i}$ to an interval that properly overlaps with $[j, \ell]$ or is a proper superset of $[j, \ell]$.

Claim 1. For all $i \in \{1, \ldots, k+1\}$ and $j \in \{1, \ldots, k\}$, $w\left[1 + \sum_{i'<i} m_{i'} n_{i'} : \sum_{i' \leq i} m_{i'} n_{i'}\right]$ does not cut $\varphi(x_j^{n_j})$.

Proof of Claim 1. We establish Claim 1 by induction on $i = 1, \ldots, k+1$. Suppose by way of a contradiction that $w[1 : m_1 n_1] = a^{m_1 n_1}$ cuts some $\varphi(x_i^{n_i})$, so that $\varphi(x_i^{n_i})$ is of the shape $a^{i_1} b^{i_2} a^{i_3} \ldots$ for some $i_1, i_2, i_3, \ldots$ with $i_1, i_2 \geq 1$. $\varphi(x_i)$ is of the shape $a^{i'_1} b^{i'_2} \ldots$, where $i'_1 = i_1$ and $i'_2 = i_2$. Note that $a^{i_1} b^{i_2} a$ cannot be a prefix of $\varphi(x_i)$; otherwise, since $n_i \geq 2$, there would be at least two occurrences of $a^{i_1} b^{i_2} a$ in w, which is false as $m_1 n_1, m_2 n_2, \ldots, m_k n_k, m_{k+1} n_{k+1}$ is strictly increasing. On the other hand, if $\varphi(x_i) = a^{i_1} b^{i_2}$, then $n_i \geq 2$ implies that $a^{i_1} b^{i_2} a^{i_1} b^{i_2}$ is a substring of w, which is also false. Thus $w[1 : m_1 n_1]$ does not cut $\varphi(x_i^{n_i})$. Proceeding inductively, assume that $w\left[1 + \sum_{i'<i} m_{i'} n_{i'} : \sum_{i' \leq i} m_{i'} n_{i'}\right]$ does not cut $\varphi(x_j^{n_j})$ for all $i \leq p$ (for some $p \leq k$) and $j \in \{1, \ldots, k\}$. Without loss of generality, suppose that $w\left[1 + \sum_{i'<p+1} m_{i'} n_{i'} : \sum_{i' \leq p+1} m_{i'} n_{i'}\right] = a^{m_{p+1} n_{p+1}}$. By the induction hypothesis, if $w\left[1 + \sum_{i'<p+1} m_{i'} n_{i'} : \sum_{i' \leq p+1} m_{i'} n_{i'}\right]$ cuts some $\varphi(x_j^{n_j})$, then I_φ must map $\left[1 + \sum_{i'<j} n_{i'} : \sum_{i' \leq j} n_{i'}\right]$ to an interval $[\ell_1, \ell_2]$ such that $1 + \sum_{i'<p+1} m_{i'} n_{i'} \leq \ell_1 \leq \sum_{i' \leq p+1} m_{i'} n_{i'} < \ell_2$, so that $\varphi(x_j^{n_j})$ is of the shape $a^{j_1} b^{j_2} a^{j_3} \ldots$, where $j_1, j_2 \geq 1$. By applying an argument similar to that for the base case, this would give a contradiction. ∎ (Claim 1)

According to Claim 1, for every factor $w\left[1 + \sum_{i'<i} m_{i'} n_{i'} : \sum_{i' \leq i} m_{i'} n_{i'}\right]$ of w, there is at least one j such that I_φ maps the interval of positions of π occupied by $x_j^{n_j}$ to a subinterval of $\left[1 + \sum_{i'<i} m_{i'} n_{i'} : \sum_{i' \leq i} m_{i'} n_{i'}\right]$. But i ranges from 1 to $k+1$ while there are only k distinct factors of π of the shape $x_j^{n_j}$, a contradiction. The rest of the proof proceeds as in the earlier proof sketch of Theorem 10(2). ∎

Appendix L. Proof of Theorem 12

Theorem 12.

1. $\text{TD}(\text{R}\Pi^1) = 3$.
2. For all $z \geq 2$ (including $z = \infty$), $\text{TD}(\text{R}\Pi^z) \geq 5$.
3. For all $z \geq 7$ (including $z = \infty$), $\text{TD}(\text{R}\Pi^z) = 5$.

Proof 1. To see that $\text{TD}(\text{R}\Pi^1) \geq 3$, note that $\text{R}\Pi^1$ contains all constant patterns and the pattern x_1. To distinguish a non-constant pattern other than x_1 from all constant patterns, at least two positive examples are needed. To distinguish it from x_1, at least one negative example is needed. Thus $\text{TD}(\text{R}\Pi^1) \geq 3$. It remains to show that every pattern in $\text{R}\Pi^1$ has a teaching set of size no larger than 3. To this end, note that patterns in $\text{R}\Pi^1$ can be normalized to either a^n or $a^{n-1}x_1$ for some $n \geq 1$. The constant pattern a^n is the only pattern in $\text{R}\Pi^1$ that is consistent with $\{(a^n, +), (a^{n+1}, -), (a^{n-1}, -)\}$.

As a teaching set for the pattern $a^{n-1}x_1$, where $n \geq 2$, one may use $\{(a^{n-1}, +), (a^n, +), (a^{n-2}, -)\}$. In case $n = 1$, i.e., for the pattern x_1, the set $\{(\varepsilon, +)\}$ suffices.

2. This part of the proof is very similar to a corresponding proof for non-erasing languages, see (Gao et al., 2016, Theorem 15). Let $z = |\Sigma| \geq 2$. Consider the pattern $\pi = ax_1bx_2a$. We claim that $\text{TD}(\pi, \text{R}\Pi^z) \geq 5$. In particular, we show that any teaching set for π w.r.t. $\text{R}\Pi^z$ contains at least two positive and three negative examples. Two positive examples are needed to distinguish π from all constant patterns. To see that three negative examples are needed, we provide three patterns $\pi_1, \pi_2, \pi_3 \in \text{R}\Pi^z$ that generate pairwise different languages such that $L(\pi_i) \cap L(\pi_j) = L(\pi)$ for $1 \leq i < j \leq 3$. Then each negative example for π rules out at most one of the three patterns π_1, π_2, π_3, so that any teaching set for π w.r.t. $\text{R}\Pi^z$ must contain at least three negative examples. The following three patterns satisfy the required conditions: $\pi_1 = x_1bx_2a$, $\pi_2 = ax_1bx_2$, and $\pi_3 = ax_1a$.

3. This result is immediate from 2. and the following sequence of lemmas.

Lemma L.1 *Let $z = |\Sigma| \geq 7$ and $n \geq 2$. Let π be any regular pattern of the shape $\pi = X_1c_1X_2c_2 \dots X_{n-1}c_{n-1}X_n$ for some $c_1, c_2, \dots, c_{n-1} \in \Sigma^+$ and $X_1, X_2, \dots, X_n \in X^+$. Then TD$(\pi, R\Pi^z) \leq 3$. In particular, π has a teaching set of size three w.r.t. $R\Pi^z$ that contains two positively labelled examples that neither start nor end with the same letter.*

Lemma L.2 *Let $z = |\Sigma| \geq 2$ and π be a regular pattern that starts and ends with a block of variables. Let T be a teaching set for π w.r.t. $R\Pi^z$ such that T contains two positively labelled examples that neither start nor end with the same letter. Let $c_1, c_2 \in \Sigma^+$. Then the following hold:*

1. *TD$(c_1\pi, R\Pi^z) \leq 1 + |T|$ and TD$(\pi c_1, R\Pi^z) \leq 1 + |T|$,*
2. *TD$(c_1\pi c_2, R\Pi^z) \leq 2 + |T|$.*

Lemma L.3 *Let $z = |\Sigma| \geq 2$. Let $c \in \Sigma^+$ and $X_1 \in X^+$ be regular patterns. Then TD$(c, R\Pi^z) =$ TD$(X_1, R\Pi^z) = 2$.*

The proofs of these lemmas are very similar (but with a few important differences) to the corresponding proofs for the non-erasing regular pattern languages; see (Gao et al., 2016, Lemmas 26 and 28). First, note that any regular pattern of the shape $X_1d_1X_2 \dots d_{h-1}X_h$, where $X_1, X_2, \dots, X_h \in X^+$, is equivalent to a regular pattern in which any two distinct variables are separated by a constant block. Every regular pattern can thus be expressed in a *canonical* form $c_1x_1c_2x_2 \dots x_{n-1}c_n$, where $c_1, c_{n-1} \in \Sigma^*$ and $c_2, \dots, c_{n-2} \in \Sigma^+$. Throughout this proof, it is assumed that every regular pattern is expressed in its canonical form. We introduce the following notation for this proof. Let $c \in \Sigma^+$. If $|\Sigma| \geq 3$ and a is a letter that differs from $c[1]$ and $c[n]$, then we define

$$\hat{c} = c^{\vdash}ac^{\dashv} \text{ for } c^{\vdash} = c[1] \dots c[|c| - 1] \text{ and } c^{\dashv} = c[2] \dots c[|c|] \ . \tag{10}$$

The notation $\hat{c}$ does not make the choice of a explicit but this choice will always be clear from the context.

Proof of Lemma L.1. Let $\prec$ be a linear order on Σ, where $z = |\Sigma| \geq 7$. Let $m = |\pi|$. For each $i \in [1, n-1]$, let i' be the maximum index less than i such that $c_{i'} \neq c_i$ (if no such index exists then set $i' = i$) and let i'' be the minimum index greater than i such that $c_{i''} \neq c_i$ (if no such index exists then set $i'' = i$). Let a_i be the least (w.r.t. $\prec$) letter in Σ such that a_i is different from the first and last symbols of any member of $\{c_{i'}, c_i, c_{i''}\}$. Define the strings α, β and γ as follows.

$$\begin{aligned}
\alpha &= \pi(\varepsilon) = c_1 c_2 \ldots c_{n-1}, \\
\beta &= \underbrace{a_1^m c_1 a_1^m}\,\underbrace{a_2^m c_2 a_2^m} \ldots \underbrace{a_{n-1}^m c_{n-1} a_{n-1}^m}, \\
\gamma &= \underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\,\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m} \ldots \underbrace{a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m} \\
&\quad \ldots \underbrace{a_{n-2}^m \hat{c}_{n-2} a_{n-2}^{2m} a_{n-1}^m c_{n-1} a_{n-1}^m a_{n-2}^m w_{n-2} a_{n-2}^m}\,\underbrace{a_{n-1}^m \hat{c}_{n-1} a_{n-1}^m},
\end{aligned}$$

where, for each $i \in [1, n-2]$,

$$w_i = \begin{cases} c_i & \text{if } c_i \neq c_{i+1}; \\ \varepsilon & \text{if } c_i = c_{i+1}. \end{cases}$$

Note that α and β neither start nor end with the same letter. We shall show that $T = \{(\alpha, +), (\beta, +), (\gamma, -)\}$ is a teaching set for π wrt. $\mathrm{R}\Pi^z$ by establishing the following claims.

Claim 1. $\alpha, \beta \in L(\pi)$ and $\gamma \notin L(\pi)$.

Claim 2. For any $\pi' \in \mathrm{R}\Pi^z$ such that $\{\alpha, \beta\} \subset L(\pi')$ and $L(\pi') \neq L(\pi)$, $\gamma \in L(\pi')$.

It is immediate from Claims 1 and 2 that for any $\pi' \in \mathrm{R}\Pi^z$ such that $L(\pi') \neq L(\pi)$, π' is inconsistent with T. This would show that T is indeed a teaching set for π w.r.t. $\mathrm{R}\Pi^z$.

Proof of Claim 1. α is obtained from π by substituting the empty string for every variable of π, and β is obtained from π by substituting a_1^m for X_1, a_{n-1}^m for X_n, and $a_{i-1}^m a_i^m$ for X_i whenever $i \in [2, n-2]$. Thus $\{\alpha, \beta\} \subset L(\pi)$. Now it is shown by induction that $\gamma \notin L(\pi)$. First, note that by construction c is not a substring of $\hat{c}$ for all $c \in \Sigma^+$. In particular, c_i is not a substring of $\hat{c}_i$ for all $i \in [1, n-1]$. Furthermore, suppose c_i were a proper substring of c_{i+1}. Then $w_i = c_i$ and c_{i+1} cannot be a substring of c_i. Combining the last two facts with the requirements on a_{i+1} and a_{i+2}, it follows that $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m$ does not contain a substring of the shape $s_1 c_1 s_2 c_2 s_3$ for any $s_1, s_2, s_3 \in \Sigma^*$. Similarly, if c_i is not a proper substring of c_{i+1}, then the definitions of w_i, a_{i+1} and a_{i+2} again imply that $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m$ does not contain a substring of the shape $s_1 c_1 s_2 c_2 s_3$ for any $s_1, s_2, s_3 \in \Sigma^*$. Assume inductively that $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m \ldots a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m$ does not contain a substring of the shape $s_1 c_1 s_2 c_2 \ldots s_{i+1} c_{i+1} s_{i+2}$ for any $s_1, s_2, \ldots, s_{i+2} \in \Sigma^*$. By the definition of a_i^m, no prefix of c_{i+1} is a suffix of $a_1^m \hat{c}_1 a_1^{2m} a_2^m\, c_2 a_2^m\, a_1^m w_1 a_1^m \ldots a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m$. Consequently, as c_{i+1} is not a substring of $a_{i+1}^m \hat{c}_{i+1} a_{i+1}^m$ and $a_{i+1}^{2m} a_{i+2}^m c_{i+2} a_{i+2}^m a_{i+1}^m w_{i+1} a_{i+1}^m$ does not contain a substring of the shape $s_1 c_{i+1} s_2 c_{i+2} s_3$ for any $s_1, s_2, s_3 \in \Sigma^*$, one has that $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m \ldots a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m a_{i+1}^m \hat{c}_{i+1} a_{i+1}^{2m} a_{i+2}^m c_{i+2} a_{i+2}^m a_{i+1}^m w_{i+1} a_{i+1}^m$ cannot be expressed in the form $s_1 c_1 s_2 c_2 \ldots s_{i+2} c_{i+2} s_{i+3}$ for any $s_1, s_2, \ldots, s_{i+3} \in \Sigma^*$. Similarly, $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m \ldots a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m a_{i+1}^m \hat{c}_{i+1} a_{i+1}^m$ cannot be expressed in the form $s_1 c_1 s_2 c_2 \ldots s_{i+1} c_{i+1} s_{i+2}$ for any $s_1, s_2, \ldots, s_{i+2} \in \Sigma^*$. It follows by induction that $\gamma \notin L(\pi)$. ▮ (Claim 1)

Proof of Claim 2. Consider any $\pi' \in \mathrm{R}\Pi^z$ such that $L(\pi') \neq L(\pi)$ and $\{\alpha, \beta\} \subset L(\pi')$. Since α and β start (as well as end) with different symbols, π' is of the shape $x_1 d_1 x_2 d_2 \ldots d_{h-1} x_h$, where $x_1, x_2, \ldots, x_h \in X$ and $d_1, d_2, \ldots, d_{h-1} \in \Sigma^+$. We claim that the following holds:

(*) Let $h : (\Sigma \cup X)^* \to \Sigma^*$ be a substitution witnessing $\beta \in L(\pi')$. Then, w.r.t. the decomposition

$$\underbrace{a_1^m}\underbrace{c_1}\underbrace{a_1^m a_2^m}\underbrace{c_2}\underbrace{a_2^m a_3^m}\ldots\underbrace{a_{i-1}^m a_i^m}\underbrace{c_i}\underbrace{a_i^m a_{i+1}^m}\ldots\underbrace{a_{n-2}^m a_{n-1}^m}\underbrace{c_{n-1}}\underbrace{a_{n-1}^m} \tag{11}$$

of β, there exists a least $i \in [1, n-1]$ such that for some $j \in [1, |c_i|]$, either (1) $h^{-1}(c_i[j])$ is a variable or (2) $h^{-1}(c_l[k])$ is a constant for all $l \in [1, n-1], k \in [1, |c_l|]$ and a variable of π' occurs between $h^{-1}(c_i[j])$ and $h^{-1}(c_i[j+1])$.

Suppose otherwise. Since $\alpha \in L(\pi')$ and π' contains at least one variable, $|\pi'(\varepsilon)| < m$. Thus, for each of the strings $a_1^m, a_1^m a_2^m, \ldots, a_{n-2}^m a_{n-1}^m, a_{n-1}^m$ indicated by braces in (11), h maps some variable of π' to at least one position in each of these strings. As $\pi' \neq \pi$, π' is in canonical form, and no variable of π' occurs between $h^{-1}(c_i[j])$ and $h^{-1}(c_i[j+1])$ for all $i \in [1, \ldots, n-1]$ and $j \in [1, |c_i|]$, $\pi'(\varepsilon)$ must be of the shape $s_1 c_1 s_2 c_2 \ldots s_{n-1}\ c_{n-1} s_n$, where $s_1, s_2, \ldots, s_n \in \Sigma^*$ and at least one s_i is nonempty. This contradicts the fact that $\alpha \in L(\pi')$ and $|\alpha| < |\pi'(\varepsilon)|$. Now let $i \in [1, n-1]$ be the least number that satisfies (*), and let $j \in [1, |c_i|]$ be the least number for which either $h^{-1}(c_i[j])$ is a variable or $h^{-1}(c_l[k])$ is a constant for all $l \in [1, n-1]$ and $k \in [1, |c_l|]$ and a variable of π' occurs between $h^{-1}(c_i[j])$ and $h^{-1}(c_i[j])$. (Note that we are referring to the specific occurrence of c_i in β indicated by the sequence of braces in the decomposition (11).) We shall define a substitution $\varphi : (\Sigma \cup X)^* \to \Sigma^*$ such that $\varphi(\pi') = \gamma$. In order to define φ, we will use the decomposition (11) of β; for each prefix $a_1^m c_1 a_1^m \ldots a_k^m c_k$ of β, φ will map $h^{-1}(a_1^m c_1 a_1^m \ldots a_k^m c_k)$ to a prefix ω of γ. (In what follows, the specific occurrence of ω in γ will be given w.r.t. the decomposition (12) of γ below.)

$$\begin{gathered}\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_i^m \hat{c}_i a_i^{2m} a_{i+1}^m c_{i+1} a_{i+1}^m a_i^m w_i a_i^m}\ldots \\ \underbrace{a_{n-2}^m \hat{c}_{n-2} a_{n-2}^{2m} a_{n-1}^m c_{n-1} a_{n-1}^m a_{n-2}^m w_{n-2} a_{n-2}^m}\underbrace{a_{n-1}^m \hat{c}_{n-1} a_{n-1}^m},\end{gathered} \tag{12}$$

Assume that $i \in [2, n-2]$. (The cases $i = 1$ and $i = n-1$ can be handled in a very similar way.) Consider the decomposition (11) of β. We first map $h^{-1}(a_1^m c_1)$ to $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1$ if $c_1 \neq c_2$, and to $a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2$ if $c_1 = c_2$. To construct such a map, note that since $|\pi'| \leq m$ and π' is not a constant pattern, there is a least position p_1 of π' occupied by a variable x_1 such that h maps x_1 to some substring of a_1^m (the first occurrence of a_1^m in the decomposition (11)). If $c_1 \neq c_2$, so that $w_1 = c_1$, then one can define $\varphi(x_1)$ to be an extension of $h(x_1)$ so that $\varphi(x_1)$ covers the substring $v \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m v'$ for some suffix v of a_1^m starting at the first position in a_1^m that h maps x_1 to and some prefix v' of a_1^m ending at the last position in a_1^m that h maps x_1 to. Letting $a_1^m = v'v''$ and $a_1^m = v'''v$ for some $v'', v''' \in \Sigma^*$, one can then define $\varphi(h^{-1}(v''w_1)) = h(h^{-1}(v''w_1)) = v''c_1$ and $\varphi(h^{-1}(v''')) = h(h^{-1}(v''')) = v'''$. If $c_1 = c_2$, so that $a_1 = a_2$, then φ can be defined so that it extends $h(x_1)$ to cover the substring $v \hat{c}_1 a_1^{2m} u$, where v is defined as above and u is the prefix of a_2^m ending at the last position in a_1^m ($= a_2^m$) that h maps x_1 to. Letting $a_2^m = uu'$ for some $u' \in \Sigma^*$, one then defines $\varphi(h^{-1}(u'c_2)) = h(h^{-1}(u'c_1)) = u'c_1$ and $\varphi(h^{-1}(v''')) = h(h^{-1}(v''')) = v'''$.

Inductively, assume that for all $k < j$, where $j < i$, φ maps $h^{-1}(a_1^m c_1 a_1^m \ldots a_k^m c_k)$ to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_k^m \hat{c}_k a_k^{2m} a_{k+1}^m c_{k+1} a_{k+1}^m a_k^m w_k}$$

if $c_k \neq c_{k+1}$, or to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_k^m \hat{c}_k a_k^{2m} a_{k+1}^m c_{k+1}}$$

if $c_k = c_{k+1}$. We now define the image of $h^{-1}(a_1^m c_1 a_1^m \ldots a_{j-1}^m c_{j-1} a_{j-1}^m a_j^m c_j)$ under φ.

Case (i): $c_{j-1} \neq c_j$. Again, since $|\pi'| \leq m$ and π' contains at least one variable, there is a least position p' such that $\pi'[p']$ is a variable x' and h maps x' to some substring of a_j^m (where the specific occurrence of a_j^m in β being referred to is indicated by braces below).

$$a_1^m c_1 a_1^m \ldots a_{j-1}^m c_{j-1} a_{j-1}^m \underbrace{a_j^m} c_j$$

Let p'' be the position of π' that h maps to the first position of the substring a_{j-1}^m whose occurrence in β is indicated by braces below.

$$a_1^m c_1 a_1^m \ldots a_{j-1}^m c_{j-1} \underbrace{a_{j-1}^m} a_j^m c_j$$

For every symbol s of π' between the $(p'')^{th}$ position and the $(p'-1)^{st}$ position inclusive, define $\varphi(s) = h(s)$. If $c_j \neq c_{j+1}$, then $\varphi(x')$ can be defined as an extension of $h(x')$ so that $\varphi(x')$ covers the substring $v_1 \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1} a_{j+1}^m v_2$ for some suffix v_1 of a_j^m starting at the first position in a_j^m that h maps x' to and some prefix v_2 of a_j^m ending at the last position in a_j^m that h maps x' to. If $c_j = c_{j+1}$ (so that $a_{j+1} = a_j$), then $\varphi(x')$ can be defined as an extension of $h(x')$ so that $\varphi(x')$ covers the substring $w_1 \hat{c}_j a_j^{2m} w_2$, where w_1 is the suffix of a_j^m starting at the first position in a_j^m that h maps x' to and w_2 is the prefix of a_{j+1}^m ending at the last position in a_j^m $(= a_{j+1}^m)$ that h maps x' to. Proceeding as in the case $j = 1$, one can then extend the definition of φ so that φ maps $h^{-1}(a_1^m c_1 a_1^m \ldots a_j^m c_j)$ to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_j^m \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1} a_{j+1}^m a_j^m w_j}$$

if $c_j \neq c_{j+1}$, and to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_j^m \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1}}$$

if $c_j = c_{j+1}$.

Case (ii): $c_{j-1} = c_j$. Then $w_{j-1} = \varepsilon$. Define $p', p'' \in \mathbb{N}$ and the variable x' as in Case (i). If $c_j \neq c_{j+1}$, then $\varphi(x')$ can be defined as an extension of $h(x')$ so that $\varphi(x')$ covers the substring $v_1 a_{j-1}^{2m} a_j^m \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1} a_{j+1}^m v_2$ for some suffix v_1 of a_j^m starting at the first position in a_j^m

that h maps x' to and some prefix v_2 of a_j^m that ends at the last position in a_j^m that h maps x' to. If $c_j = c_{j+1}$ (so that $a_{j+1} = a_j$), then $\varphi(x')$ can be defined as an extension of $h(x')$ so that $\varphi(x')$ covers the substring $u_1 a_j^m a_{j-1}^{2m} a_j^m \hat{c}_j a_j^{2m} u_2$, where u_1 is the suffix of a_j^m starting at the first position in a_j^m that h maps x' to and u_2 is the prefix of a_{j+1}^m ending at the last position in a_j^m $(= a_{j+1}^m)$ that h maps to. Proceeding as in the case $j = 1$, one can then extend the definition of φ so that φ maps $h^{-1}(a_1^m c_1 a_1^m \ldots a_j^m c_j)$ to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_j^m \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1} a_{j+1}^m a_j^m w_j}$$

if $c_j \neq c_{j+1}$, and to

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_j^m \hat{c}_j a_j^{2m} a_{j+1}^m c_{j+1}}$$

if $c_j = c_{j+1}$. .

For $j = i$, φ maps the string

$$a_1^m c_1 a_1^m \ldots a_{i-1}^m c_{i-1} a_{i-1}^m a_i^m c_i$$

to the substring

$$\underbrace{a_1^m \hat{c}_1 a_1^{2m} a_2^m c_2 a_2^m a_1^m w_1 a_1^m}\underbrace{a_2^m \hat{c}_2 a_2^{2m} a_3^m c_3 a_3^m a_2^m w_2 a_2^m}\ldots\underbrace{a_i^m \hat{c}_i};$$

note that such a mapping can be defined because either $h^{-1}(c_i[j])$ (w.r.t. the decomposition (11)) is a variable for at least one $j \in [1, |c_i|]$, or $h^{-1}(c_l[k])$ is a constant for all $l \in [1, n-1]$ and $k \in [1, |c_l|]$ and π' contains a variable between $h^{-1}(c_i[j])$ and $h^{-1}(c_i[j+1])$ for some $j \in [1, |c_i|]$. To see this, first suppose there exists some q' such that q' is the least position of π' for which $\pi'[q']$ is a variable y and h maps y to some substring of c_i; now choose the least j such that h maps y to the j^{th} position of c_i. (The specific occurrence of c_i in β being referred to is indicated by braces below.)

$$a_1^m c_1 a_1^m \ldots a_{i-1}^m c_{i-1} \underbrace{a_{i-1}^m}\underbrace{a_i^m}\underbrace{c_i} \tag{13}$$

Let θ and η be strings such that $\hat{c}_i = c_i[1] \ldots c_i[j-1] \theta c_i[j] \alpha c_i[j+1] \ldots c_i[|c_i|]$. One can define $\varphi(y)$ so that $\varphi(y)$ covers the substring $\theta c_i[j] \alpha$ of $\hat{c}_i$. Now consider the following case distinction.

Case (i): $c_{i-1} \neq c_i$. Define $\varphi(h^{-1}(a_{i-1}^m a_i^m c_i[1] \ldots c_i[j-1])) = a_{i-1}^m a_i^m c_i[1] \ldots c_i[j-1]$ (as a prefix of $a_{i-1}^m a_i^m \hat{c}_i$) and $\varphi(h^{-1}(c_i[j+1] \ldots c_i[|c_i|])) = c_i[j+1] \ldots c_i[|c_i|]$ (as a suffix of $a_{i-1}^m a_i^m \hat{c}_i$).

Case (ii): $c_{i-1} = c_i$. Then $w_{i-1} = \varepsilon$ and $a_{i-1} = a_i$. There is a least position r of a_{i-1}^m (where a_{i-1}^m is indicated by braces in (13)) such that for some variable z of π', h maps z to the r^{th} position of a_{i-1}^m. $\varphi(z)$ can be defined as an extension of $h(z)$ so that $\varphi(z)$ covers $u_1 a_{i-1}^m u_2$, where u_1 is the suffix of a_i^m that starts at the r^{th} position of a_i^m and u_2 is the prefix of a_{i-1}^m $(= a_i^m)$ that ends at the last position of a_{i-1}^m that h maps z to. Letting $a_1^m = u_3 u_1 = u_2 u_4$ for some $u_3, u_4 \in \Sigma^*$, define $\varphi(h^{-1}(u_3)) = u_3$ and $\varphi(h^{-1}(u_4 a_i^m c_i[1] \ldots c_i[j-1])) = u_4 a_i^m c_i[1] \ldots c_i[j-1]$ (as a prefix of $u_4 a_i^m \hat{c}_i$) and $\varphi(h^{-1}(c_i[j+1] \ldots c_i[|c_i|])) = c_i[j+1] \ldots c_i[|c_i|]$ (as a suffix of $u_4 a_i^m \hat{c}_i$).

Now suppose that $h^{-1}(c_l[k])$ is a constant for all $l \in [1, n-1]$ and $k \in [1, |c_l|]$ and π' contains a variable z between $h^{-1}(c_i[j])$ and $h^{-1}(c_i[j+1])$ for some $j \in [1, |c_i|]$. The definition of $\varphi(h^{-1}(a_1^m c_1 a_1^m \ldots a_{i-1}^m c_{i-1} a_{i-1}^m a_i^m c_i))$ here is very similar to that in the previous case. Let θ' be the string such that $\hat{c}_i = c_i[1] \ldots c_i[j]\theta' c_i[j+1] \ldots c_i[|c_i|]$. One can define $\varphi(z)$ so that $\varphi(z)$ covers the substring θ' of $\hat{c}_i$. Further, one defines $\varphi(h^{-1}(a_{i-1}^m a_i^m c_i[1] \ldots c_i[j]))$ and $\varphi(h^{-1}(c_i[j+1] \ldots c_i[|c_i|]))$ according to a case distinction similar to that in the previous case.

By applying an argument similar to that in the preceding paragraph, one can extend the definition of φ to $h^{-1}(a_1^m c_1 a_1^m \ldots a_{j-1}^m c_{j-1} a_{j-1}^m a_j^m c_j a_j^m)$ for all $j \in [1, n-1]$. ▮ (Claim 2)

This establishes that T is a teaching set for π w.r.t. $\mathrm{R}\Pi^z$. ▮ (Lemma L.1)

Proof of Lemma L.2. We prove that $\mathrm{TD}(c_1\pi c_2, \mathrm{R}\Pi^z) \leq 2 + |T|$; the remaining cases can be proved similarly. We follow the proof of (Gao et al., 2016, Lemma 28) (the analogue of Lemma L.2 for the class of non-erasing pattern languages). Suppose T is a teaching set for π w.r.t. $\mathrm{R}\Pi^z$ containing at least two positively labelled examples $(w_1, +), (w_2, +)$ that neither start nor end with the same letter. Let $T' = \{(c_1 w c_2, +) : (w, +) \in T\} \cup \{(c_1 v c_2, -) : (v, -) \in T\} \cup \{(\hat{c}_1 w_1 c_2, -), (c_1 w_1 \hat{c}_2, -)\}$. Let $\pi' = d_1 \rho d_2$ be a regular pattern that is consistent with T', where ρ starts and ends with variables and $d_1, d_2 \in \Sigma^*$. Since $(c_1 w_1 c_2, +), (c_1 w_2 c_2, +) \in L(\pi')$ and w_1, w_2 both start as well as end with different symbols, d_1 is a prefix of c_1 and d_2 is a suffix of c_2. We argue that d_1 is in fact equal to c_1. Let $\varphi : (\Sigma \cup X)^* \to \Sigma^*$ be a substitution witnessing $c_1 w_1 c_2 \in L(\pi')$. If $d_1 = c_1[1] \ldots c_1[k]$ for some $k < |c_1|$, then $v = \hat{c}_1 w_1 c_2 \in L(\pi')$: one can map the variable x_1 in π' occurring just after d_1 to $c_1[k+1] \ldots c_1[|c_1| - 1]ac[2] \ldots c_1[2] \ldots c_1[|c_1|]$ (where $a \notin \{c_1[1], c_1[|c_1|]\}$), and for each position j of v after $\hat{c}_1$, one maps $\varphi^{-1}(v[j])$ (which may be equal to x_1) to $v[j]$. This contradicts the fact that π' is consistent with T'. A similar argument shows that if d_2 were a proper suffix of c_2, then $v' = c_1 w_1 \hat{c}_2 \in L(\pi')$, a contradiction. Thus $\pi' = c_1 \rho c_2$. Furthermore, note that for all $u \in \Sigma^*$ and $l \in \{+, -\}$, $\pi' = c_1 \rho c_2$ is consistent with $(c_1 u c_2, l)$ iff ρ is consistent with (u, l). Hence if T is a teaching set for π w.r.t. $\mathrm{R}\Pi^z$, then T' is a teaching set for $c_1 \pi c_2$ w.r.t. $\mathrm{R}\Pi^z$. ▮ (Lemma L.2)

Proof of Lemma L.3. Let $c \in \Sigma^+$ and $X_1 \in X^+$ for some regular pattern X_1. Fix distinct $a, b \in \Sigma$. One may directly verify that $\{(c, +), (c^2, -)\}$ is a teaching set for c w.r.t. $\mathrm{R}\Pi^z$ while $\{(a, +), (b, +)\}$ is a teaching set for X_1 w.r.t. $\mathrm{R}\Pi^z$. Furthermore, $\mathrm{TD}(c, \mathrm{R}\Pi^z) \geq 2$ because a single positive example is consistent with X_1 while a single negative example $(v, -)$ for some $v \in \Sigma^*$ is consistent with c' for any $c' \in \Sigma^* \setminus \{c, v\}$. Also, $\mathrm{TD}(X_1, \mathrm{R}\Pi^z) \geq 2$ because a single positive example $(w, +)$ is consistent with w while every teaching set for X_1 contains only positive examples. ▮ (Lemma L.3)

■

Appendix M. Proof of Theorem 14

Theorem 14. Let $z \in \mathbb{N} \cup \{\infty\}$.

1. No recursive teaching sequence for $1\Pi^z$ exists.

2. If $z \geq 2$, then no recursive teaching sequence for $\mathrm{NC}\Pi^z$ exists.

3. $\mathrm{RTD}(\mathrm{NC}\Pi^1) = \infty$.

Proof 1. Suppose there is a recursive teaching sequence $\mathcal{S} = ((S_0, d_0), (S_1, d_1), \ldots)$ for $1\Pi^z$. Let $a \in \Sigma$ and let $\pi_0 = a$ be a constant pattern. Let $i_0 \in \mathbb{N}$ such that $\pi_0 \in S_{i_0}$. Let $d = \max\{d_i \mid i \leq i_0\}$. In particular, every pattern in $S_0 \cup \ldots \cup S_{i_0}$ has a recursive teaching set of size at most d w.r.t. $\mathcal{S}$.

Let T_0 be a recursive teaching set for π_0 with respect to $\mathcal{S}$. Now choose $d+i_0+1$ distinct primes $p_1 < p_2 < \ldots < p_{d+i_0+1}$ such that p_1 is strictly greater than all the lengths of the strings in T_0. Let $P_0 = \{p_1, \ldots, p_{d+i_0+1}\}$. Define a pattern $\pi_1 = ax^{q_0}$, where $q_0 = \prod_{p \in P_0} p$. Any positive example in T_0 must be for the string a, which is in $L(\pi_1)$. As q_0 is strictly greater than all the lengths of the strings in T_0, π_1 cannot generate any negative example in T_0. Hence π_1 is consistent with T_0 and thus must belong to some L_{i_1} where $i_1 < i_0$.

Let T_1 be a recursive teaching set for π_1 with respect to $\mathcal{S}$. Let P_1 be any $(d+i_0)$-subset of P_0. Define a pattern $\pi_2 = ax^{q_1}$, where $q_1 = \prod_{p \in P_1} p$. Then π_2 is consistent with T_1 and thus must belong to some L_{i_2} where $i_2 < i_1$.

Iterating this argument, we obtain a $(d+1)$-subset P of P_0 such that $ax^q \in S_0$ for $q = \prod_{p \in P} p$. Now observe that $\mathrm{TD}(ax^q, 1\Pi^z) \geq d+1$, which contradicts the statement that every pattern in $S_0 \cup \ldots \cup S_{i_0}$ has a recursive teaching set of size at most d w.r.t. $\mathcal{S}$. Therefore, no recursive teaching sequence for $1\Pi^z$ exists.

2. Note that by the proof of Theorem 10(2), all non-cross patterns π not equivalent to the pattern x have infinite teaching dimension w.r.t. the class of all non-cross patterns π' such that $L(\pi') \neq L(x)$. Thus there is no teaching sequence for the class C of all non-cross pattern languages $L(\pi)$ such that $L(\pi) \neq L(x)$ because the first concept to be taught in any such sequence already has infinite teaching dimension w.r.t. C.

3. This follows immediately from the fact that the RTD of the class $\{\{\boldsymbol{v}^\top \boldsymbol{x} : \boldsymbol{x} \in \mathbb{N}_0^n\} : 0 \neq \boldsymbol{v} \in \mathbb{N}_0^n \wedge n \geq 1\}$ is infinite (Gao et al., 2015, Corollary 16). ■

Appendix N. Proof of Theorem 15

Theorem 15. Let $z \in \mathbb{N} \cup \{\infty\}$. If $z \neq 2$, then $\mathrm{RTD}(\mathrm{R}\Pi^z) = 2$.

Proof For any z, one obtains $\mathrm{RTD}(\mathrm{R}\Pi^z) \geq 2$ from the obvious fact that no regular pattern other than x_1 has a teaching set of size one w.r.t. $\mathrm{R}\Pi^z \setminus \{x_1\}$. Now we only need to show the existence of a teaching sequence $\mathcal{S}$ of order 2 for $\mathrm{R}\Pi^z$.

Let us first consider the case $z = 1$ and $\Sigma = \{a\}$. Then any regular pattern can be normalised to either a^n or $a^{n-1}x_1$ for some $n \geq 1$. The teaching sequence $\mathcal{S}$ lists patterns in increasing order of the number of constant symbols. The pattern $a^{n-1}x_1$ uses $\{(a^{n-1}, +), (a^n, +)\}$ as a recursive teaching set w.r.t. $\mathcal{S}$, while a^n uses $\{(a^n, +), (a^{n+1}, -)\}$.

Now let $z \geq 3$. Then any regular pattern can be normalised to a form like $c_1x_1c_2 \ldots c_nx_nc_{n+1}$ where $n \geq 0$, $c_1, c_{n+1} \in \Sigma^*$ and $c_i \in \Sigma^+$ for $2 \leq i \leq n$. The teaching sequence $\mathcal{S}$ lists patterns in increasing order of the number of constants. Patterns with the same number of constants are listed in decreasing order of the number of variables. Let $\pi = c_1x_1c_2 \ldots c_nx_nc_{n+1}$ be a normalised regular pattern as above. Let $w \in \Sigma^+$ be the string generated by π when replacing any variable x_i with a symbol $a_i \in \Sigma$ such that a_i is different from the last symbol of c_i (if $c_i \neq \varepsilon$) and the first symbol of c_{i+1} (if $c_{i+1} \neq \varepsilon$). Since $z \geq 3$, this is possible. We then claim that $T = \{(\pi(\varepsilon), +), (w, +)\}$ is a recursive teaching set for π w.r.t. $\mathcal{S}$.

By choice of the sequence $\mathcal{S}$, the set T needs to distinguish π only from (i) those regular patterns that have more than $|\pi(\varepsilon)|$ constants, as well as (ii) those with exactly $|\pi(\varepsilon)|$ constants and at most n variables. (i) is achieved by the example $(\pi(\varepsilon), +)$, which now rules out all patterns π' for which $\pi'(\varepsilon) \neq \pi(\varepsilon) = c_1 \ldots c_{n+1}$. Note that $w = c_1a_1c_2 \ldots c_na_nc_{n+1}$.

Suppose a regular pattern π' with $\pi'(\varepsilon) = c_1 \dots c_{n+1}$ generates w, where π' has at most n variables. Let φ be a substitution that maps π' to w. If φ did not map the first occurrence of c_1 in π' to the first occurrence of c_1 in w, then φ would have to map the first occurrence of c_1 in π' to a substring in w that starts at least two positions later than the first occurrence of c_1 in w (otherwise c_1 would have to end in a_1). Two positions after the first occurrence of c_1 in w, the first occurrence of c_2 after c_1 in w begins. Repeating this argument, for $2 \leq i \leq n$ (if $c_{n+1} = \varepsilon$) or for $2 \leq i \leq n+1$ (if $c_{n+1} \neq \varepsilon$), φ maps the first occurrence of c_i after c_{i-1} in π' to at least two positions to the right of the first occurrence of c_i after c_{i-1} in w (Note that since a_i differs from the first letter of c_{i+1}, c_{i+1} cannot start at a_i.) This would require $|\varphi(\pi')| > |w|$ in contradiction to $\varphi(\pi') = w$. Thus φ maps the first occurrence of c_1 in π' to the first occurrence of c_1 in w, and, inductively, for $2 \leq i \leq n+1$, φ maps the first occurrence of c_i after c_{i-1} in π' to the first occurrence of c_i after c_{i-1} in w. This is only possible if $\pi' = \pi$. ■