Codes Defined by Multiple Sets of Trajectories [3]

Michael Domaratzki [a,*,1], Kai Salomaa [b,2]

[a] *Jodrey School of Computer Science, Acadia University, Wolfville, NS Canada B4P 2R6*

[b] *School of Computing, Queen's University, Kingston, ON Canada K7L 3N6*

Abstract

We investigate the use of shuffle on trajectories to model certain classes of languages arising in the theory of codes. In particular, for each finite set of sets of trajectories, which we call a hyperset of trajectories, we define a class of languages induced by that hyperset of trajectories. We investigate properties of hypersets of trajectories and the associated classes of languages, including the problem of decidability of membership and the problem of equivalence of hypersets of trajectories.

Key words: shuffle on trajectories, deletion along trajectories, theory of codes

1 Introduction

Shuffle on trajectories was introduced as a framework for modelling language operations which act by inserting the letters of one word into another word in a sequential manner [26]. This is accomplished by modelling a particular operation by a set of trajectories, a language over a binary alphabet. Since the introduction of shuffle on trajectories, there has been substantial interest in

* Corresponding Author.
Email addresses: `mike.domaratzki@acadiau.ca` (Michael Domaratzki), `ksalomaa@cs.queensu.ca` (Kai Salomaa).

[1] Research supported in part by a grant from the Natural Sciences and Engineering Research Council of Canada.

[2] Research supported in part by the Natural Sciences and Engineering Research Council of Canada grant OGP0147224.

[3] A preliminary version of this paper appeared at AFL 2005 [9].

several different related areas, including algebraic properties of shuffle on trajectories [25,27], language equations defined by shuffle on trajectories [2,8,20], generalizations [6,19,24], and applications to areas such as noisy channels [19] and DNA computing [18]. The first author has presented a survey of recent results on trajectories [7].

A large amount of work in the literature has considered the problems of examining code-like properties of languages in several different general frameworks [12–15,17,30]. Recently, trajectories have also been employed to model classes of languages related to codes [4] and the natural binary relation defined by a set of trajectories has also been examined [16,5]. The use of trajectories is a natural way to study certain classes of languages related to codes, which we call *T-codes*.

However, there exist natural classes of languages studied in connection to the theory of codes which are not T-codes. Classes and their associated binary relations studied by Day and Shyr [1], Fan *et al.* [10], Ito *et al.* [11], Long [21], Long *et al.* [22,23], Shyr [29], Yu [31] and the first author [3] are instead defined by a binary relation dependent on multiple sets of trajectories.

In this paper, we study the classes of languages which are naturally defined by multiple sets of trajectories. We examine decidability and find that, surprisingly, the natural, uniform membership problem is undecidable for regular languages and regular sets of trajectories. We also examine the equivalence problem for the classes of languages defined by multiple sets of trajectories. Decidability of this problem remains open.

2 Definitions

Let Σ be a finite set of symbols, called *letters*. Then Σ^* is the set of all finite sequences of letters from Σ, which are called *words*. The empty word ϵ is the empty sequence of letters. The *length* of a word $w = w_1 w_2 \cdots w_n \in \Sigma^*$, where $w_i \in \Sigma$, is n, and is denoted $|w|$. For any $w \in \Sigma^*$ and $a \in \Sigma$, we denote by $|w|_a$ the number of occurrences of a in w. A *language* L is any subset of Σ^*. By abuse of notation, we represent the singleton language $\{w\}$ by w.

If $L_1, L_2 \subseteq \Sigma^*$, then the concatenation of L_1 and L_2 is denoted $L_1 L_2$, and is given by $L_1 L_2 = \{xy \ : \ x \in L_1, y \in L_2\}$. Further, if L_i is a language for $1 \leq i \leq n$, then by $\prod_{i=1}^{n} L_i$ we mean the language $L_1 L_2 \cdots L_n$.

A *morphism* $h : \Delta^* \to \Sigma^*$ is any function satisfying $h(xy) = h(x)h(y)$ for all $x, y \in \Delta^*$. For additional background in formal languages and automata theory, please see Rozenberg and Salomaa [28]. We denote the finite, regular, context-free and recursive languages by FIN,REG,CF and REC, respectively.

The shuffle on trajectories operation is a method for specifying the ways in which two input words may be merged, while preserving the order of symbols in each word. Each trajectory $t \in \{0,1\}^*$ with $|t|_0 = n$ and $|t|_1 = m$ specifies one particular way in which we can form the shuffle on trajectories of two words of length n (as the left operand) and m (as the right operand). The word resulting from the shuffle along t will have length $n+m$, with a letter from the left input word in position i if the i-th symbol of t is 0, and a letter from the right input word in position i if the i-th symbol of t is 1.

We now give the definition of shuffle on trajectories, originally due to Mateescu *et al.* [26]. Shuffle on trajectories is defined by first defining the shuffle of two words x and y over an alphabet Σ on a trajectory t, a word over $\{0,1\}$. We denote the shuffle of x and y on trajectory t by $x \sqcup\!\sqcup_t y$.

If $x = ax'$, $y = by'$ (with $a, b \in \Sigma$) and $et \in \{0,1\}^*$ (with $e \in \{0,1\}$), then

$$x \sqcup\!\sqcup_{et} y = \begin{cases} a(x' \sqcup\!\sqcup_t by') & \text{if } e = 0; \\ b(ax' \sqcup\!\sqcup_t y') & \text{if } e = 1. \end{cases}$$

If $x = ax'$ ($a \in \Sigma$), $y = \epsilon$ and $et \in \{0,1\}^*$ ($e \in \{0,1\}$), then

$$x \sqcup\!\sqcup_{et} \epsilon = \begin{cases} a(x' \sqcup\!\sqcup_t \epsilon) & \text{if } e = 0; \\ \emptyset & \text{otherwise.} \end{cases}$$

If $x = \epsilon$, $y = by'$ ($b \in \Sigma$) and $et \in \{0,1\}^*$ ($e \in \{0,1\}$), then

$$\epsilon \sqcup\!\sqcup_{et} y = \begin{cases} b(\epsilon \sqcup\!\sqcup_{t'} y') & \text{if } e = 1; \\ \emptyset & \text{otherwise.} \end{cases}$$

We let $x \sqcup\!\sqcup_\epsilon y = \emptyset$ if $\{x, y\} \neq \{\epsilon\}$. Finally, if $x = y = \epsilon$, then $\epsilon \sqcup\!\sqcup_t \epsilon = \epsilon$ if $t = \epsilon$ and $\emptyset$ otherwise.

It is not difficult to see that if $t = \prod_{i=1}^n 0^{j_i} 1^{k_i}$ for some $n \geq 0$ and $j_i, k_i \geq 0$ for all $1 \leq i \leq n$, then we have that

$$\begin{aligned} x \sqcup\!\sqcup_t y = \{ \prod_{i=1}^n x_i y_i \ \ : \\ & x = \prod_{i=1}^n x_i, y = \prod_{i=1}^n y_i, \text{ with } |x_i| = j_i, |y_i| = k_i \text{ for all } 1 \leq i \leq n \} \end{aligned}$$

if $|x| = |t|_0$ and $|y| = |t|_1$, and $x \sqcup\!\sqcup_t y = \emptyset$ if $|x| \neq |t|_0$ or $|y| \neq |t|_1$.

We extend shuffle on trajectories to *sets* $T \subseteq \{0,1\}^*$ *of trajectories* as follows:

$$x \sqcup\!\sqcup_T y = \bigcup_{t \in T} x \sqcup\!\sqcup_t y.$$

Further, for $L_1, L_2 \subseteq \Sigma^*$, we define

$$L_1 \sqcup\!\sqcup_T L_2 = \bigcup_{\substack{x \in L_1 \\ y \in L_2}} x \sqcup\!\sqcup_T y.$$

Thus, for example, it is not hard to see that if $T = 0^*1^*$, then $L_1 \sqcup\!\sqcup_T L_2 = L_1 L_2$ (the usual concatenation operation) while if $T = 0^*1^*0^*$, $L_1 \sqcup\!\sqcup_T L_2 = L_1 \leftarrow L_2$, the insertion operation, defined by $x \leftarrow y = \{x_1 y x_2 \ : \ x = x_1 x_2\}$.

We also require the definition of the natural binary relation defined by a set of trajectories. For all $T \subseteq \{0,1\}^*$, let ω_T be the binary relation on Σ^* defined by

$$x \, \omega_T \, y \iff y \in x \sqcup\!\sqcup_T \Sigma^*$$

for all $x, y \in \Sigma^*$. The relation ω_T has previously been studied by the first author [5], as well as by Kadrie *et al.* [16] for infinite strings. Consider the following examples:

(i) If $T = 0^*1^*$, then ω_T is the prefix relation, i.e., $x \, \omega_T \, y$ if and only if there exists $z \in \Sigma^*$ such that $y = xz$.
(ii) If $T = 1^*0^*1^*$, then ω_T is the factor or subword relation, i.e., $x \, \omega_T \, y$ if and only if there exists $w_1, w_2 \in \Sigma^*$ such that $y = w_1 x w_2$.
(iii) If $T = \{0,1\}^*$, then ω_T is the embedding relation (or substring relation).

Let $\mathcal{P}_T(\Sigma)$ be the set of non-empty subsets of Σ^+ which are anti-chains under ω_T. Equivalently, a non-empty language $L \subseteq \Sigma^+$ is in $\mathcal{P}_T(\Sigma)$ if and only if the equality $L \cap (L \sqcup\!\sqcup_T \Sigma^+) = \emptyset$ holds. If $L \in \mathcal{P}_T(\Sigma)$ we say that L is a T-*code*. Investigation of the classes $\mathcal{P}_T(\Sigma)$ was undertaken by the first author [4]. As an example, if $T = 0^*1^*$, then $\mathcal{P}_T(\Sigma)$ is the set of prefix codes over Σ. If $T = \{0,1\}^*$, the associated class $\mathcal{P}_T(\Sigma)$ is known as the set of hypercodes (see, e.g., Shyr [29, Sect. 5.2]).

Let $\mathbf{T} \subseteq 2^{\{0,1\}^*}$. We call such a set of sets of trajectories $\mathbf{T}$ a *hyperset of trajectories*; such a hyperset of trajectories is always assumed to be a finite set of sets of trajectories. Define $\omega_{\mathbf{T}}$ as

$$x \, \omega_{\mathbf{T}} \, y \iff \bigwedge_{T \in \mathbf{T}} x \, \omega_T \, y.$$

That is, $x \, \omega_{\mathbf{T}} \, y$ if and only if $x \, \omega_T \, y$ for all $T \in \mathbf{T}$.

The class $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ is defined as follows: for all non-empty languages $L \subseteq \Sigma^+$, $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ if and only if L is an anti-chain under $\omega_{\mathbf{T}}$. That is, for all $x, y \in L$, if $x \, \omega_{\mathbf{T}} \, y$, then $x = y$.

The definition of $\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ is motivated by the interest in the class $\mathcal{P}^{(\wedge)}_{\mathbf{T}_{\rm ps}}(\Sigma)$ for $\mathbf{T}_{\rm ps} = \{0^*1^*, 1^*0^*\}$. Note that $x \,\omega_{\mathbf{T}_{\rm ps}}\, y$ implies that x is both a prefix and a suffix of y: $x \,\omega_{0^*1^*}\, y$ implies that $y = xz$ for some $z \in \Sigma^*$ and similarly $x \,\omega_{1^*0^*}\, y$ implies that $y = wx$ for some $w \in \Sigma^*$. We refer the reader to Jürgensen and Konstantinidis [12, pp. 550–551] for references and a discussion of $\mathcal{P}^{(\wedge)}_{\mathbf{T}_{\rm ps}}(\Sigma)$.

We also define a second class of languages, indexed by an integer m, which has often been considered in conjunction with $\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ for particular $\mathbf{T}$. For all $m \geq 0$, let $\mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ be defined as follows: for all non-empty languages $L \subseteq \Sigma^+$, $L \in \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ if and only if for all $L' \subseteq L$ with $|L'| \leq m$, $L' \in \cup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma)$. That is, $L \in \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ if, for all $L' \subseteq L$ of size at most m, there exists $T \in \mathbf{T}$ such that $L' \in \mathcal{P}_T(\Sigma)$. We note that the definition of the class $\mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ is an $(m+1)$-dependence system, in the terminology of Jürgensen and Konstantinidis [12].

The class $\mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ has been studied for the following hypersets of trajectories:

(i) $\mathbf{T}_{\rm ps} = \{0^*1^*, 0^*1^*\}$. The class $\mathcal{P}^{(m)}_{\mathbf{T}_{\rm ps}}(\Sigma)$ is known as the class of *m-prefix-suffix codes* (or m-ps-codes). See Ito *et al.* [11] for details;
(ii) $\mathbf{T}_{\rm io} = \{0^*1^*0^*, 1^*0^*1^*\}$ [23,3]. The class $\mathcal{P}^{(m)}_{\mathbf{T}_{\rm io}}(\Sigma)$ is known as the class of *m-infix-outfix codes*;
(iii) $\mathbf{T}_{k-{\rm io}} = \{(1^*0^*)^k 1^*, (0^*1^*)^k 0^*\}$ and $\mathbf{T}_{k-{\rm ps}} = \{(0^*1^*)^k, (1^*0^*)^k\}$ for $k \geq 1$. The class $\mathcal{P}^{(m)}_{\mathbf{T}_{k-{\rm io}}}(\Sigma)$ (resp., $\mathcal{P}^{(m)}_{\mathbf{T}_{k-{\rm ps}}}(\Sigma)$) is known as the class of *m-k-infix-outfix codes* (resp., *m-k-prefix-suffix codes*). For results on these classes, see Long *et al.* [22, Sect. 4] or Long [21, Sect. 2.3]).

3 Properties

We begin with some elementary properties of the classes $\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ and $\mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$. First, we note that hypersets of size one give us precisely T-codes:

Lemma 1 *For all $T \subseteq \{0,1\}^*$, if $\mathbf{T} = \{T\}$, then $\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma) = \mathcal{P}_T(\Sigma)$.*

Further, containment between hypersets of trajectories implies containment between the associated classes of languages.

Lemma 2 *Let $\mathbf{S}, \mathbf{T} \subseteq 2^{\{0,1\}^*}$ with $\mathbf{S} \subseteq \mathbf{T}$. Then $\mathcal{P}^{(\wedge)}_{\mathbf{S}}(\Sigma) \subseteq \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$. Further, for all $m \geq 1$, $\mathcal{P}^{(m)}_{\mathbf{S}}(\Sigma) \subseteq \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$.*

PROOF. Let $L \in \mathcal{P}^{(\wedge)}_{\mathbf{S}}(\Sigma)$. Assume there exist $x, y \in L$ such that $x \,\omega_{\mathbf{T}}\, y$. Then $x \,\omega_T\, y$ for all $T \in \mathbf{T}$. In particular, as $\mathbf{T} \supseteq \mathbf{S}$, $x \,\omega_S\, y$ for all $S \in \mathbf{S}$. Thus,

$x\,\omega_{\mathbf{S}}\,y$ and $x = y$, as $L \in \mathcal{P}^{(\wedge)}_{\mathbf{S}}(\Sigma)$. Thus, $L \in \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$. This establishes the first statement.

For the second statement, let $m \geq 1$. Let $L \in \mathcal{P}^{(m)}_{\mathbf{S}}(\Sigma)$ and $L' \subseteq L$ with $|L'| \leq m$ be arbitrary. Then there exists $S \in \mathbf{S} \subseteq \mathbf{T}$ such that $L' \in \mathcal{P}_S(\Sigma)$. Thus, as S is indeed in $\mathbf{T}$, we have that $L \in \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$. $\square$

For all $\mathbf{T} \subseteq 2^{\{0,1\}^*}$, let $\mathbf{min}(\mathbf{T})$ be defined by

$$\mathbf{min}(\mathbf{T}) = \{T \in \mathbf{T} \;:\; \forall T' \in \mathbf{T}, T \not\supset T'\}.$$

That is, $\mathbf{min}(\mathbf{T})$ is the set of minimal elements of $\mathbf{T}$, as a subset of the lattice $2^{\{0,1\}^*}$ ordered by inclusion.

Lemma 3 *Let $\mathbf{T} \subseteq 2^{\{0,1\}^*}$. Then the following equalities holds for all alphabets Σ:*

$$\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma) = \mathcal{P}^{(\wedge)}_{\mathbf{min}(\mathbf{T})}(\Sigma),$$
$$\mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma) = \mathcal{P}^{(m)}_{\mathbf{min}(\mathbf{T})}(\Sigma) \quad \forall m.$$

PROOF. Let us establish the first equality. As $\mathbf{min}(\mathbf{T}) \subseteq \mathbf{T}$, $\mathcal{P}^{(\wedge)}_{\mathbf{min}(\mathbf{T})}(\Sigma) \subseteq \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ by Lemma 2. Thus, it remains to show the reverse inclusion. Let $L \notin \mathcal{P}^{(\wedge)}_{\mathbf{min}(\mathbf{T})}(\Sigma)$. Let $x, y \in L$ with $x \neq y$ be such that $x \omega_{\mathbf{min}(\mathbf{T})} y$. In particular, $x\,\omega_T\,y$ for all $T \in \mathbf{min}(\mathbf{T})$.

Let $T_0 \in \mathbf{T} - \mathbf{min}(\mathbf{T})$. Then there exists $T_1 \in \mathbf{min}(\mathbf{T})$ such that $T_1 \subseteq T_0$. As $T_1 \in \mathbf{min}(\mathbf{T})$, $x\,\omega_{T_1}\,y$, i.e., $y \in x \sqcup_{T_1} \Sigma^*$. As $T_1 \subseteq T_0$, $y \in x \sqcup_{T_0} \Sigma^*$. Thus, $x\,\omega_{T_0}\,y$. In this way, we have that $x\,\omega_T\,y$ for all $T \in \mathbf{T}$. Thus, $x\,\omega_{\mathbf{T}}\,y$ and $L \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$.

Now, we establish the second equality. Let $m \geq 1$. Once again, as $\mathbf{min}(\mathbf{T}) \subseteq \mathbf{T}$, $\mathcal{P}^{(m)}_{\mathbf{min}(\mathbf{T})}(\Sigma) \subseteq \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$ by Lemma 2. Let $L \in \mathcal{P}^{(m)}_{\mathbf{T}}(\Sigma)$. Then for all $L' \subseteq L$, with $|L'| \leq m$, there exists $T \in \mathbf{T}$ such that $L' \in \mathcal{P}_T(\Sigma)$.

Consider now an arbitrary $L' \subseteq L$ with $|L'| \leq m$. Then $L' \in \mathcal{P}_T(\Sigma)$ for some $T \in \mathbf{T}$. If $T \in \mathbf{min}(\mathbf{T})$, then we are done. Otherwise, if $T \in \mathbf{T} - \mathbf{min}(\mathbf{T})$, then there exists $T' \in \mathbf{min}(\mathbf{T})$ such that $T \supset T'$. Assume, contrary to what we want to prove, that $L' \notin \mathcal{P}_{T'}(\Sigma)$. Then there exists $x, y \in L'$ and $t \in T'$ such that $y \in x \sqcup_{t'} \Sigma^+$. But $t \in T$ as well, so $L \notin \mathcal{P}_T(\Sigma)$, a contradiction. Thus, $L \in \mathcal{P}_{T'}(\Sigma)$.

Therefore, for arbitrary $L' \subseteq L$ with $|L'| \leq m$, there exists $T \in \mathbf{min}(\mathbf{T})$ such that $L' \in \mathcal{P}_{T'}(\Sigma)$, as required. $\square$

We say that $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ is *minimal* if $\mathbf{T} = \mathbf{min}(\mathbf{T})$.

4 Relationships Between Classes

In this section, we investigate various relationships between the classes $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ and $\mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma)$. We begin with a result which motivates our introduction of the classes $\mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma)$—we prove that $\mathcal{P}_{\mathbf{T}}^{(2)}(\Sigma)$ is always exactly $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$.

Lemma 4 *Let $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ be a hyperset of trajectories. Then*

$$\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma) = \mathcal{P}_{\mathbf{T}}^{(2)}(\Sigma).$$

PROOF. Let $L \in \mathcal{P}_{\mathbf{T}}^{(2)}(\Sigma)$. Let $u, v \in L$ and suppose $u \omega_{\mathbf{T}} v$. Therefore, $u \omega_T v$ for all $T \in \mathbf{T}$. As $L \in \mathcal{P}_{\mathbf{T}}^{(2)}(\Sigma)$, $\{u, v\} \in \mathcal{P}_T(\Sigma)$ for some $T \in \mathbf{T}$. Thus, $u = v$. We conclude that L is an anti-chain under $\omega_{\mathbf{T}}$ and $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$.

For the reverse inclusion, let $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$. Assume $L \notin \mathcal{P}_{\mathbf{T}}^{(2)}(\Sigma)$. Then there exists $\{u, v\} \subseteq L$ $(u \neq v)$ such that $\{u, v\} \notin \cup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma)$. Thus, $\{u, v\} \notin \mathcal{P}_T(\Sigma)$ for all $T \in \mathbf{T}$. We have that $u \, \omega_T \, v$ or $v \, \omega_T \, u$ for all $T \in \mathbf{T}$. Assume that there exist $T_1, T_2 \in \mathbf{T}$ such that $u \, \omega_{T_1} \, v$ and $v \, \omega_{T_2} \, u$. Then $u \, \omega_{T_1} \, v$ implies $|v| \geq |u|$ and $v \, \omega_{T_2} \, u$ implies $|u| \geq |v|$. Thus $|u| = |v|$. As $v \in u \sqcup\!\sqcup_{T_1} \Sigma^*$, this implies that $u = v$. Thus, we must have without loss of generality that $u \, \omega_T \, v$ holds for all $T \in \mathbf{T}$. Thus, $u \omega_{\mathbf{T}} v$, and consequently, $u = v$, as $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$. $\square$

Lemma 4 was previously observed for, e.g., the case $T_{\text{ps}} = \{0^*1^*, 1^*0^*\}$; see Ito *et al.* [11]. The following equations detail the hierarchies induced by varying m in $\mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma)$, and their collapse. These equations, which hold for all $T \subseteq 2^{\{0,1\}^*}$, can be proven using dependency theory [12] (in the following, $|\mathbf{T}|$ is the cardinality of $\mathbf{T}$ as a subset of $2^{\{0,1\}^*}$):

$$\mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma) \supseteq \mathcal{P}_{\mathbf{T}}^{(m+1)}(\Sigma) \quad \forall m \geq 0; \tag{1}$$

$$\mathcal{P}_{\mathbf{T}}^{(2|\mathbf{T}|)}(\Sigma) = \mathcal{P}_{\mathbf{T}}^{(2|\mathbf{T}|+i)}(\Sigma) = \bigcup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma) \quad \forall i \geq 0. \tag{2}$$

See, e.g., Ito *et al.* [11, Cor. 3.2] for (2) in the particular case of $T_{\text{ps}} = \{0^*1^*, 1^*0^*\}$.

We now consider whether there exists a hyperset of trajectories for which the m-th and $m+1$-st levels of the above hierarchy are distinct for all $m \geq 1$. The answer is yes, but clearly by (2), the size of $\mathbf{T}$ must depend on m.

Lemma 5 *Let $m \geq 1$. There exists $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ such that $\mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma) \neq \mathcal{P}_{\mathbf{T}}^{(m+1)}(\Sigma)$.*

PROOF. Let $m \geq 1$ and $t_i = 1^{i-1}01^{m-i+1}$ for all $1 \leq i \leq m$. We let $\mathbf{T} = \{\{t_i\}\}_{i=1}^m$. Let $L = \{t_i\}_{i=1}^m \cup \{0\}$. Note that $|L| = m+1$. It is now straightforward to verify that $L \in \mathcal{P}_{\mathbf{T}}^{(m)}(\Sigma) - \mathcal{P}_{\mathbf{T}}^{(m+1)}(\Sigma)$. $\square$

The following inclusion holds between $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ and the union of the associated class of T-codes:

Lemma 6 *For all $\mathbf{T} \subseteq 2^{\{0,1\}^*}$,*

$$\bigcup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma) \subseteq \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma).$$

PROOF. Let $L \in \cup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma)$. In particular, let $T_0 \in \mathbf{T}$ be such that $L \in \mathcal{P}_{T_0}(\Sigma)$. Let $x, y \in L$ and assume, contrary to what we want to prove, that $x \,\omega_{\mathbf{T}}\, y$. Therefore, $x \,\omega_T\, y$ for all $T \in \mathbf{T}$ and in particular, $x \,\omega_{T_0}\, y$. Thus $x = y$ as $L \in \mathcal{P}_{T_0}(\Sigma)$. Therefore, L is an anti-chain under $\omega_{\mathbf{T}}$, and $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$. $\square$

Under certain additional conditions, the previous inclusion is proper if we allow the alphabet size to grow in relation to $|\mathbf{T}|$. We say that $T \subseteq \{0,1\}^*$ is *ST-strict* if $0^* + 1^* \subseteq T$. We say that $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ is ST-strict if T is ST-strict for all $T \in \mathbf{T}$.

We say that $\mathbf{T}$ is *fully incomparable* if, for all $T \in \mathbf{T}$, we have that

$$T - \bigcup_{S \in \mathbf{T} - \{T\}} S \neq \emptyset.$$

Note that full incomparability implies minimality.

Lemma 7 *Let $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ with $|\mathbf{T}| \geq 2$ be fully incomparable and ST-strict. Then*

$$\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma) - \bigcup_{T \in \mathbf{T}} \mathcal{P}_T(\Sigma) \neq \emptyset$$

for all Σ with $|\Sigma| \geq |\mathbf{T}| + 1$.

PROOF. As $\mathbf{T}$ is fully incomparable, for all $T \in \mathbf{T}$, there exists

$$t_T \in T - \bigcup_{\substack{T' \in \mathbf{T} \\ T' \neq T}} T'.$$

Let $\alpha_T = |t_T|_0$. As each T is ST-strict, $\alpha_T \neq 0$. For all $T \in \mathbf{T}$, let a_T be a distinct letter. Let $h_T : \{0,1\}^* \to \{a_T, b\}$ be given by $h_T(0) = a_T$ and $h_T(1) = b$. Let $\Sigma_{\mathbf{T}} = \{a_T \ : \ T \in \mathbf{T}\} \cup \{b\}$. We now define $L \subseteq \Sigma_{\mathbf{T}}^+$ as follows:

$$L = \bigcup_{T \in \mathbf{T}} \{a_T^{\alpha_T}, h_T(t_T)\}.$$

Note that $h_T(t_T) \neq a_T^{\alpha_T}$, as $\mathbf{T}$ is ST-strict.

We now establish that L satisfies the conditions of the lemma: let $T \in \mathbf{T}$. Then as $h_T(t_T) \in a_T^{\alpha_T} \sqcup_{t_T} b^+$, $L \notin \mathcal{P}_T(\Sigma_{\mathbf{T}})$.

Assume that $L \notin \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma_{\mathbf{T}})$. Then there exist $u, v \in L$ such that $u \omega_{\mathbf{T}} v$. That is, $u \, \omega_T \, v$ for all $T \in \mathbf{T}$. As each pair $\{a_T^{\alpha_T}, h_T(t_T)\} \subseteq \{a_T, b\}^+$ and $|h_T(t_T)|_b \neq 0$ for all $T \in \mathbf{T}$, we must have that $u = a_{T_0}^{\alpha_{T_0}}$ and $v = h_{T_0}(t_{T_0})$ for some $T_0 \in \mathbf{T}$. Therefore, $h_{T_0}(t_{T_0}) \in a_{T_0}^{\alpha_{T_0}} \sqcup_T \Sigma_{\mathbf{T}}^+$ for all $T \in \mathbf{T}$. By the definition of $\sqcup_T$, this implies that $t_{T_0} \in T$ for all $T \in \mathbf{T}$, a contradiction to our choice of t_{T_0}. Therefore, $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma_{\mathbf{T}})$. The result now follows, as $|\Sigma_{\mathbf{T}}| = |\mathbf{T}| + 1$. $\square$

In general, some increase in alphabet size is necessary, as is shown in the following lemma:

Lemma 8 *There exists $\mathbf{T} = \{T_1, T_2, T_3\}$ which is fully incomparable and ST-strict such that*

$$\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\{a,b\}) = \mathcal{P}_{T_1}(\{a,b\}) \cup \mathcal{P}_{T_2}(\{a,b\}) \cup \mathcal{P}_{T_3}(\{a,b\}).$$

PROOF. Let $T = \{01, 10, 001, 010, 011, 100, 101, 110\}$, $T_i = T \cup \{0^{4-i}1^i\} \cup 0^* \cup 1^*$ for $1 \leq i \leq 3$ and $\mathbf{T} = \{T_1, T_2, T_3\}$.

Let $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\{a,b\}) - \cup_{i=1}^3 \mathcal{P}_{T_i}(\{a,b\})$. As $L \notin \mathcal{P}_{T_i}(\{a,b\})$, there exist $x_i, y_i \in L$ with $x_i \neq y_i$ and $x_i, y_i \neq \epsilon$ such that $x_i \, \omega_{T_i} \, y_i$ for all $1 \leq i \leq 3$. Note that $|x_i| < |y_i|$. Further, for all $1 \leq i \leq 3$, $|y_i| \leq 4$.

Assume that $|y_i| < 4$ for some $1 \leq i \leq 3$. Then let $t_i \in T_i$ and $\alpha_i \in \{a,b\}^+$ be such that $y_i \in x_i \sqcup_{t_i} \alpha_i$. As $|t_i| = |y_i| < 4$, $t_i \in T$. Thus, $x_i \, \omega_{T_j} \, y_i$ for all $1 \leq j \leq 3$, whereby $x_i \, \omega_{\mathbf{T}} \, y_i$, which contradicts that $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\{a,b\})$.

Thus, $|y_i| = 4$ for all $1 \leq i \leq 3$. Let $t_i = 0^{4-i}1^i$. We must have that $y_i \in x_i \sqcup_{t_i} \{a,b\}^+$ for all $1 \leq i \leq 3$. Thus, by definition of $\sqcup_T$, $|x_i| = 4 - i$.

Consider $i = 3$. Then $|x_3| = 1$. Without loss of generality, let $x_3 = a$. For $i = 2$, we have that $|x_2| = 2$. Assume that $x_2 \neq b^2$. Then $x_2 \in x_3 \sqcup_T \{a,b\}$, by our choice of T. Thus, $x_3 \, \omega_{T_i} \, x_2$ for all $1 \leq i \leq 3$ and therefore $x_3 \, \omega_{\mathbf{T}} \, x_2$, again a contradiction. Thus, $x_2 = b^2$.

We now turn to $i = 1$, and x_1, which satisfies $|x_1| = 3$. We again see that if $x_1 \neq b^3$ then $x_3 \, \omega_{\mathbf{T}} \, x_1$. Thus, $x_1 = b^3$. But now $x_1 \in x_2 \sqcup_T b$. Therefore, $x_2 \, \omega_{\mathbf{T}} \, x_1$. Thus $L \notin \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\{a, b\})$. This is a contradiction. Therefore $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\{a, b\}) = \cup_{i=1}^{3} \mathcal{P}_{T_i}(\{a, b\})$. $\square$

5 Decidability

We now consider decidability questions. For T-codes (i.e., for $\mathcal{P}_T(\Sigma)$), given a context-free set of trajectories T and $L \in \text{REG}$, it is decidable whether $L \in \mathcal{P}_T(\Sigma)$ [4]. However, we see here that the situation is more complicated for hypersets of trajectories.

Given a hyperset of trajectories in which one of the sets of trajectories is finite, the membership problem for recursive languages is decidable.

Lemma 9 *Let $\mathbf{T} = \{T_1, \ldots, T_n\}$ where $T_i \in \text{FIN}$ for some i with $1 \leq i \leq n$, and $T_i \in \text{REC}$ for all $1 \leq i \leq n$. Given a recursive language $L \in \Sigma^*$, the problem $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$? is decidable.*

PROOF. Let $1 \leq i \leq n$ be chosen so that T_i is finite. Let $m = \max\{|t| : t \in T_i\}$. Consider that if $x \, \omega_{\mathbf{T}} \, y$, then in particular $y \in x \sqcup_{T_i} \Sigma^*$, and thus $|y| \leq m$. Thus, in order to test if $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$, it suffices to test those words in $L^{(m)} = L \cap \Sigma^{\leq m}$. Note that as L is recursive, $L^{(m)}$ is an effective finite set.

Let $x, y \in L^{(m)}$ be arbitrary. Let $r = |y|$. As T_j is recursive for all $1 \leq j \leq n$, we can effectively determine all $t_j \in T_j$ with $|t_j| = r$ and test if $y \in x \sqcup_{t_j} \Sigma^*$. Thus, we can determine if $x \, \omega_{T_j} \, y$ for all $1 \leq j \leq n$, i.e., whether $x \, \omega_{\mathbf{T}} \, y$. Thus, it is decidable whether $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$. $\square$

We now consider the decidability of membership in $\mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ where $\mathbf{T}$ consists entirely of regular languages. The positive decidability of membership in $\mathcal{P}_{\mathbf{T}_{\text{ps}}}^{(\wedge)}(\Sigma)$ for regular languages (see Ito *et al.* [11] or Jürgensen *et al.* [13]) relies intrinsically on the nature of the members of $\mathbf{T}_{\text{ps}}$ (recall $\mathbf{T}_{\text{ps}} = \{1^*0^*, 0^*1^*\}$). The corresponding positive decidability problem for $\mathbf{T}_{\text{io}}$ also relies on the nature of the sets of trajectories involved [3] ($\mathbf{T}_{\text{io}} = \{1^*0^*1^*, 0^*1^*0^*\}$). Kari *et al.* [18, Thm. 4.7] have resolved the decidability of a somewhat similar decision problem for two sets of trajectories in their framework of *bond-free property*. However, their approach is not applicable to our formalism. We recall a particular case of their result, translated into our framework (the most general result presented by Kari *et al.* involves involutions of interest in DNA research, but is also not applicable to our situation):

Theorem 10 *Let $\mathbf{T} = \{T_1, T_2\}$ be a hyperset of trajectories where $T_i \in$ REG for $i = 1, 2$. Given a regular language $R \subseteq \Sigma^*$, it is decidable whether there exist $w_1, w_2 \in R$, and $w \in \Sigma^+$ such that $w \omega_{T_i} w_i$ for $i = 1, 2$ and either $w \neq w_1$ or $w \neq w_2$.*

In fact, we have the following rather surprising undecidability result:

Theorem 11 *Given $\boldsymbol{T} = \{T_1, T_2\}$, where $T_i \in$ REG, for $i = 1, 2$ and a regular language R it is undecidable whether or not $R \in \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$.*

PROOF. Let $I = (u_1, \ldots u_k; v_1, \ldots v_k)$, $u_i, v_i \in \Omega^+$, $i = 1, \ldots, k$ be an arbitrary instance of the Post correspondence problem.

Let $d : \Omega^* \to \Omega^*$ be the morphism defined by the condition $d(a) = aa$ for all $a \in \Omega$. Denote

$$M = \max\{|d(u_i)|, |d(v_i)| \mid 1 \leq i \leq k\}.$$

Further, we choose an injective mapping $f : \{1, \ldots, k\} \to \mathbb{N}$ such that

$$(\forall i, j \in \{1, \ldots k\}) \;\; i \neq j \text{ implies } |f(i) - f(j)| > 2M \tag{3}$$

and

$$(\forall i, j \in \{1, \ldots k\}) \;\; \frac{2}{3} f(i) < f(j) - M. \tag{4}$$

The conditions (3) and (4) can always be satisfied by choosing $f(i)$ to be sufficiently large. The reader can verify that for all $k, M \geq 1$, the function $f(i) = 6kM + 3iM$ satisfies (3) and (4).

Denote $\Sigma = \Omega \cup \{\#, \$\}$, where $\#, \$ \notin \Omega$, and define the regular language $R \subseteq \Sigma^*$ by

$$R = (\{d(u_i)\$^{f(i)} d(v_i)\$^{f(i)} \mid 1 \leq i \leq k\}^+ \cdot \#) + (\Omega^2)^*.$$

We define the hyperset of trajectories $\mathbf{T} = \{T_1, T_2\}$ where

$$T_1 = \{0^{|d(u_i)|} 1^{2f(i) + |d(v_i)|} \mid 1 \leq i \leq k\}^+ \cdot 1,$$
$$T_2 = \{1^{|d(u_i)| + f(i)} 0^{|d(v_i)|} 1^{f(i)} \mid 1 \leq i \leq k\}^+ \cdot 1.$$

Note that T_1, T_2 are both regular sets of trajectories. We claim that $R \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ if and only if the PCP instance I has a solution.

Firstly, assume that I has a solution $(i_1, \ldots, i_n)$, $n \geq 1$. This implies that the equality $d(u_{i_1}) \cdots d(u_{i_n}) = d(v_{i_1}) \cdots d(v_{i_n})$ holds. Let us denote this word by x. Since $x \in (\Omega^2)^*$, we have $x \in R$. We observe that

$$d(u_{i_1})\$^{f(i_1)} d(v_{i_1})\$^{f(i_1)} \cdots d(u_{i_n})\$^{f(i_n)} d(v_{i_n})\$^{f(i_n)}\# \in d(u_{i_1}) \cdots d(u_{i_n}) \sqcup\!\sqcup_{T_1} \Sigma^*,$$

and,

$$d(u_{i_1})\$^{f(i_1)}d(v_{i_1})\$^{f(i_1)}\cdots d(u_{i_n})\$^{f(i_n)}d(v_{i_n})\$^{f(i_n)}\# \in d(v_{i_1})\cdots d(v_{i_n}) ⧢_{T_2}\Sigma^*.$$

Since the word $d(u_{i_1})\$^{f(i_1)}d(v_{i_1})\$^{f(i_1)}\cdots d(u_{i_n})\$^{f(i_n)}d(v_{i_n})\$^{f(i_n)}\#$ is in R, this means that $R \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$.

Conversely, assume that there exist $x, y \in R$, $x \neq y$, such that

$$y \in x ⧢_{T_i}\Sigma^*, \quad i = 1, 2. \tag{5}$$

Since for any $t \in T_i$, $|t|$ is odd and $|t|_0$ is even, it follows that $|x|$ is even and $|y|$ is odd and so $x \in (\Omega^2)^*$ and $y \in \{d(u_i)\$^{f(i)}d(v_i)\$^{f(i)} \mid 1 \leq i \leq k\}^+ \cdot \#$. Denote

$$y = d(u_{i_1})\$^{f(i_1)}d(v_{i_1})\$^{f(i_1)}\cdots d(u_{i_n})\$^{f(i_n)}d(v_{i_n})\$^{f(i_n)}\#, \quad n \geq 1. \tag{6}$$

By (5) there exist $t_1 \in T_1$ and $w \in \Sigma^*$ such that $y = x ⧢_{t_1} w$. By the definition of T_1 there exist $r \geq 1$ and $1 \leq j_s \leq k$ for $1 \leq s \leq r$ such that we can write

$$t_1 = 0^{|d(u_{j_1})|}1^{2f(j_1)+|d(v_{j_1})|}0^{|d(u_{j_2})|}1^{2f(j_2)+|d(v_{j_2})|}\cdots 0^{|d(u_{j_r})|}1^{2f(j_r)+|d(v_{j_r})|}\cdot 1.$$

Let $z = 1^{2f(j_1)+|d(v_{j_1})|}$ and $t_1' \in \{0,1\}^*$ be such that $t_1 = 0^{|d(u_{j_1})|}zt_1'$. Consider that $|z| \leq 2f(j_1)+M$. By (4), $2f(j_1)+M < 3f(i)$ for all $1 \leq i \leq k$. Therefore, z cannot "produce" [4] more than two (entire) subwords $\$^{f(i)}$, $1 \leq i \leq k$, in the shuffle $x ⧢_{t_1} w$. Note that by (4) with $i = j = j_1$ we certainly have

$$\frac{1}{3}f(j_1) > M. \tag{7}$$

For the lower bound for the length of z we get

$$\begin{aligned} |z| &> 2f(j_1) = \frac{3}{2}f(j_1) + \frac{1}{2}f(j_1) \\ &> f(i) + \frac{3}{2}M + \frac{1}{2}f(j_1) \\ &> f(i) + 2M. \end{aligned}$$

for any $1 \leq i \leq k$. Above the second inequality follows by (4) and the last inequality by (7). The above means that z has to "produce" more than one subword $\$^{f(i)}$ in the shuffle $x ⧢_{t_1} w$, (that is, it has to produce at least part of a second \$-subword, since the gaps between \$-subwords is at most M). Since

[4] By "produce", we mean that for each trajectory symbol of z, a single symbol appears in the result y. Thus, we say that each of these symbols in y is produced by z.

the word x doesn't contain any symbols \$, it follows that z has to produce exactly two subwords of the form $\$^{f(i)}$.

Now, consider that the prefix of x of length $|d(u_{j_1})|$ must produce only symbols from Ω, as $x \in \Omega^*$. Therefore, z must begin by producing symbols of $d(u_{i_1})$, which is the initial block of y consisting of letters from Ω. Taken together, the above discussion implies that the prefix of w of length $|z|$ is of the form $\alpha\$^{f(i_1)}d(v_{i_1})\$^{f(i_1)}\beta$ for some $\alpha, \beta \in \Omega^*$.

Now we claim that (3) implies that $\alpha = \beta = \epsilon$ and $i_1 = j_1$. Now, consider that

$$|\alpha| + |\beta| + 2f(i_1) + |d(v_{i_1})| = 2f(j_1) + |d(v_{j_1})| \tag{8}$$

as the first is the length of the block of y "produced by" z and the second is the length of z, which are necessarily equal. Rearranging, and taking absolute values, we get

$$2|f(j_1) - f(i_1)| = ||\alpha| + |\beta| + |d(v_{i_1})| - |d(v_{j_1})||\,.$$

Now, consider the inequalities

$$0 \leq |\alpha|, |\beta| \leq M \quad \text{ and } \quad 2 \leq |d(v_{i_1})|, |d(v_{j_1})| \leq M. \tag{9}$$

The first inequality follows from the fact that as α, β are subwords of $d(u_{i_1})$ and $d(u_{i_2})$, respectively, we have $|\alpha|, |\beta| \leq M$. From (9), we get that

$$2|f(j_1) - f(i_1)| = ||\alpha| + |\beta| + |d(v_{i_1})| - |d(v_{j_1})|| \leq 3M - 2. \tag{10}$$

Assume now that $i_1 \neq j_1$. By (3), we have that $|f(j_1) - f(i_1)| > 2M$, thus, by (10), $4M < 3M - 2$. This is a contradiction and $i_1 = j_1$. Consequently, by (8), $\alpha = \beta = \epsilon$.

Continuing inductively we see that necessarily $r = n$ and $j_s = i_s$ for all $s = 1, \ldots, n$. Thus,

$$x = d(u_{i_1}) \cdots d(u_{i_n}).$$

In a completely similar way, by considering $t_2 \in T_2$ and $w' \in \Sigma^*$ such that $y = x \sqcup_{t_2} w'$ we get that

$$x = d(v_{i_1}) \cdots d(v_{i_n}).$$

This means that the PCP instance I has a solution $(i_1, \ldots, i_n)$. $\square$

We also have the following undecidability result:

Theorem 12 *There exists a fixed hyperset of trajectories* $\mathbf{T} = \{T_1, T_2\}$ *where* $T_i \in \text{REG}$ *for* $i = 1, 2$, *such that the following problem is undecidable: "Given* $L \in \text{CF}$, *is* $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$*?"*

PROOF. The problem is undecidable, e.g., for $\mathbf{T}_{\text{io}} = \{0^*1^*0^*, 1^*0^*1^*\}$ [3]. $\square$

Finally, we have the following open problem:

Open Problem 13 *For which hypersets of trajectories* $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ *is the following problem decidable: "Given* $L \in \text{REG}$*, is* $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$*?"*

It is conceivable that the question stated in Open Problem 13 could be decidable for all hypersets $\mathbf{T} = \{T_1, \ldots, T_n\}$ where $T_i \in \text{REG}$ for $1 \leq i \leq n$, in particular, if the alphabet Σ is fixed. If this is the case, by Theorem 11, given $\mathbf{T}$, the corresponding algorithm cannot be found effectively. We recall that positive decidability results for $\mathbf{T}_{\text{ps}} = \{0^*1^*, 1^*0^*\}$ [11] and $\mathbf{T}_{\text{io}} = \{0^*1^*0^*, 1^*0^*1^*\}$ [3] are known.

6 Equivalence and Slices

We now focus on the equivalence problem for hypersets of trajectories. In particular, given $\mathbf{T_1}, \mathbf{T_2} \subseteq 2^{\{0,1\}^*}$, we say that $\mathbf{T_1}$ and $\mathbf{T_2}$ are $\wedge$-equivalent with respect to Σ if $\mathcal{P}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) = \mathcal{P}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$. We simply say that $\mathbf{T_1}, \mathbf{T_2}$ are $\wedge$-equivalent if they are $\wedge$-equivalent with respect to every finite alphabet Σ. We use the notation $\mathbf{T}_1 \equiv_\wedge \mathbf{T}_2$ to indicate that $\mathbf{T}_1, \mathbf{T}_2$ are $\wedge$-equivalent.

Similarly, we say that $\mathbf{T_1}, \mathbf{T_2}$ are m-equivalent with respect to Σ if $\mathcal{P}_{\mathbf{T}_1}^{(m)}(\Sigma) = \mathcal{P}_{\mathbf{T}_2}^{(m)}(\Sigma)$. Again, we say that $\mathbf{T}_1, \mathbf{T}_2$ are m-equivalent if they are m-equivalent with respect to every finite alphabet Σ. We use the notation $\mathbf{T}_1 \equiv_m \mathbf{T}_2$ to indicate that $\mathbf{T}_1, \mathbf{T}_2$ are m-equivalent.

We first consider $\wedge$-equivalence. Note that $\wedge$-equivalence and 2-equivalence are identical conditions, by Lemma 4. Let $m, n \geq 0$. Given $\mathbf{T} \subseteq 2^{\{0,1\}^*}$, the (m,n)-*slice of* $\mathbf{T}$, denoted $\Upsilon(\mathbf{T}; m, n)$, is defined by

$$\Upsilon(\mathbf{T}; m, n) = \mathbf{min}(\{\Psi^{-1}([m,n]) \cap T \ : \ T \in \mathbf{T}\}).$$

Here, Ψ is the Parikh mapping, defined by $\Psi(w) = (|w|_0, |w|_1)$, and extended to languages as $\Psi(L) = \cup_{w \in L} \Psi(w)$. Note that $\emptyset \in \Upsilon(\mathbf{T}; m, n)$ is possible, in which case $\Upsilon(\mathbf{T}; m, n) = \{\emptyset\}$.

Lemma 14 *Let* $m, n \geq 0$*. Let* $\mathbf{T} \subseteq 2^{\{0,1\}^*}$ *satisfy* $\Upsilon(\mathbf{T}; m, n) \neq \{\emptyset\}$*. Then for all* Σ *and all* $a \in \Sigma$*,* $a^m \; \omega_{\mathbf{T}} \; a^{m+n}$*.*

PROOF. Let $\mathbf{T} = \{T_1, \ldots, T_n\}$. Then as $\{\emptyset\} \neq \Upsilon(\mathbf{T}; m, n)$, we have $\emptyset \notin \Upsilon(\mathbf{T}; m, n)$ by the minimality of $\Upsilon(\mathbf{T}; m, n)$. Therefore, for all $1 \leq i \leq n$, there exists $t_i \in T_i$ such that $\Psi(t_i) = [m, n]$. It is easy to verify that $a^{m+n} \in a^m \sqcup\!\sqcup_{t_i} a^n$. Thus, $a^m \omega_{T_i} a^{m+n}$ for all $1 \leq i \leq n$, which establishes the lemma. $\square$

We now show that if $\mathbf{T}_1, \mathbf{T}_2$ always have equal (m, n)-slices, then they are $\wedge$-equivalent. This motivates the name slices: if we take a hyperset of trajectories and separate it into its slices, then recombining the slices arbitrarily always yields a $\wedge$-equivalent hyperset of trajectories.

Lemma 15 *Let $\mathbf{T}_1, \mathbf{T}_2 \subseteq 2^{\{0,1\}^*}$. For all Σ, $\Upsilon(\mathbf{T}_1; m, n)$ and $\Upsilon(\mathbf{T}_2; m, n)$ are $\wedge$-equivalent for all $m, n \geq 0$ if and only if $\mathbf{T}_1, \mathbf{T_2}$ are $\wedge$-equivalent.*

PROOF. In what follows, let $\Upsilon_i = \Upsilon(\mathbf{T}_i; m, n)$ for $i = 1, 2$. ($\Rightarrow$): Assume without loss of generality that $L \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma) - \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$. As $L \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$, there exist $x, y \in L$ such that $x\, \omega_{\mathbf{T}_2}\, y$. Let $|x| = m$ and $|y| = m + n$.

As in the proof of Lemma 4, we see that $x\, \omega_{\Upsilon_2}\, y$. Thus, $\{x, y\} \notin \mathcal{P}^{(\wedge)}_{\Upsilon_2}(\Sigma) = \mathcal{P}^{(\wedge)}_{\Upsilon_1}(\Sigma)$. Thus, $x\, \omega_{\Upsilon_1}\, y$. As $L \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma)$, $x\, \omega_{\mathbf{T}_1}\, y$ does not hold. Thus, there exists $T \in \mathbf{T}_1$ such that $x\, \omega_T\, y$ does not hold. Let $S \in \Upsilon_1$ be such that $S \subseteq \Psi^{-1}(m, n) \cap T$. As $x\, \omega_{\Upsilon_1}\, y$, $x\, \omega_S\, y$ and thus $x\, \omega_T\, y$ holds as well. This is a contradiction. Therefore $\mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma) = \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$.

($\Leftarrow$): Let $m, n \geq 0$ be chosen so that Υ_1, Υ_2 are not $\wedge$-equivalent. Without loss of generality, let $L \in \mathcal{P}^{(\wedge)}_{\Upsilon_1}(\Sigma) - \mathcal{P}^{(\wedge)}_{\Upsilon_2}(\Sigma)$.

Let $x, y \in L$ be such that $x\, \omega_{\Upsilon_2}\, y$. As $|t|_0 = m$ and $|t|_1 = n$ for all $t \in \Upsilon_2$, we must have that $|x| = m$ and $|y| = m + n$. We now claim that $\{x, y\} \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma) - \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$.

Assume first that $\{x, y\} \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma)$. Then $x\, \omega_T\, y$ for all $T \in \mathbf{T}_1$. However, this implies that $x\, \omega_S\, y$ for all $S \in \Upsilon_1$. But then $L \notin \mathcal{P}^{(\wedge)}_{\Upsilon_1}(\Sigma)$, a contradiction. Thus, $\{x, y\} \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_1}(\Sigma)$.

Now, assume that $\{x, y\} \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$. Thus, $x\, \omega_{\mathbf{T}_2}\, y$ does not hold. Therefore, there exists $T \in \mathbf{T}_2$ such that $x\, \omega_T\, y$ does not hold. Now, there must exist $S \in \Upsilon_2$ such that $S \subseteq T \cap \Psi^{-1}([m, n])$. Further, as $x\, \omega_{\Upsilon_2}\, y$, $x\, \omega_S\, y$. But now $x\, \omega_T\, y$, a contradiction. Therefore, $\{x, y\} \notin \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$. This completes the proof. $\square$

We will require the following Lyndon-Schützenberger Theorem (see, e.g., Shyr [29, Lemma 1.6]):

Theorem 16 *Let $x \in \Sigma^*$ and $y, z \in \Sigma^+$ be such that $xy = zx$. Then there exist $\alpha, \beta \in \Sigma^*$ and $e \geq 0$ such that $x = (\alpha\beta)^e\alpha$, $y = \beta\alpha$ and $z = \alpha\beta$.*

In Lemma 15, $\wedge$-equivalence implies only the equivalence of slices, rather than equality. This is demonstrated in the following lemma:

Lemma 17 *Let $n \geq 1$, $\mathbf{T} = \{\{0^n1\}, \{10^n\}\}$ and $\mathbf{T}_1, \mathbf{T}_2 \subseteq (2^{\Psi^{-1}([n,1])} - \{\emptyset\})$ satisfy $\mathbf{T} \subseteq \mathbf{T}_i$ for $i = 1, 2$. Then $\mathcal{P}^{(\wedge)}_{\mathbf{T_1}}(\Sigma) = \mathcal{P}^{(\wedge)}_{\mathbf{T_2}}(\Sigma)$.*

PROOF. Assume that $\mathcal{P}^{(\wedge)}_{\mathbf{T_1}}(\Sigma) \neq \mathcal{P}^{(\wedge)}_{\mathbf{T_2}}(\Sigma)$. Without loss of generality, we let L be a language such that $L \in \mathcal{P}^{(\wedge)}_{\mathbf{T_1}}(\Sigma) - \mathcal{P}^{(\wedge)}_{\mathbf{T_2}}(\Sigma)$. As $L \notin \mathcal{P}^{(\wedge)}_{\mathbf{T_2}}(\Sigma)$, let $x, y \in L$ be such that $x \,\omega_{\mathbf{T_2}}\, y$. By choice of $\mathbf{T}_2$, $|x| = n$ and $|y| = n + 1$. Further, as $\{0^n1\}, \{10^n\} \in \Upsilon(\mathbf{T}_2, n, 1)$, we must have that $y = xa$ and $y = bx$ for some $a, b \in \Sigma$. From this, we can easily see that $a = b$, $x = a^n$ and $y = a^{n+1}$ by Theorem 16. Thus, by Lemma 14 and our choice of $\mathbf{T}_1$, $x \,\omega_{\mathbf{T_1}}\, y$. Thus, $L \notin \mathcal{P}^{(\wedge)}_{\mathbf{T_1}}(\Sigma)$, which contradicts our choice of L. $\square$

Using the same idea as in Lemma 17 we can construct much more complex hypersets of trajectories which are $\wedge$-equivalent but have unequal slices. In particular, if we allow $\mathbf{T}_1, \mathbf{T}_2$ to be the $(n, 1)$-slices of two larger hypersets of trajectories, all of whose other slices are equal, then the result still holds.

For instance, if $T_1, T_2 \subseteq \{0, 1\}^*$ are arbitrary sets of trajectories for which $\Psi(T_1) = \Psi(T_2)$ then we have that the following hypersets of trajectories are $\wedge$-equivalent:

$$\mathbf{T}_1 = \{0^*1, 10^*, T_1\},$$
$$\mathbf{T}_2 = \{0^*1, 10^*, T_2\}.$$

We note that if $T_1 \neq T_2$, we do not necessarily have $\Upsilon(\mathbf{T}_1; m, n) = \Upsilon(\mathbf{T}_2; m, n)$ for any $m, n \geq 0$.

It is clear that given two hypersets of trajectories $\mathbf{T}_1, \mathbf{T}_2$ such that the inclusions $\Psi(\mathbf{T}_1), \Psi(\mathbf{T}_2) \subseteq [m, n]$ for some $m, n \geq 0$, we can determine whether $\mathbf{T}_1 \equiv_\wedge \mathbf{T}_2$. However, the following problem is still open:

Open Problem 18 *Given $\mathbf{T}_1, \mathbf{T}_2$, each consisting of regular sets of trajectories, can we determine whether $\mathbf{T}_1$ and $\mathbf{T}_2$ are $\wedge$-equivalent?*

We can restate Open Problem 18 as follows:

Open Problem 19 *For given regular sets of trajectories $T_1, \cdots, T_k \subseteq \{0, 1\}^*$ and a regular set of trajectories $U \subseteq \{0, 1\}^*$ is it decidable whether or not there* *exist an alphabet* Σ *and* $x, y \in \Sigma^+$ *such that for all* $1 \leq i \leq k$, $y \in x \sqcup\!\sqcup_{T_i} \Sigma^+$ *but* $y \notin x \sqcup\!\sqcup_U \Sigma^+$*?*

The problem of m-equivalence of hypersets of trajectories is left as a topic for future research.

7 Maximality

The concept of maximality is key in the theory of codes. Maximal codes are codes none of whose proper supersets are also codes. This concept has been examined extensively with respect to many subclasses of codes, for instance, maximal prefix codes. In this section, we consider maximality with respect to the classes we have considered in this paper.

We define $\mathcal{M}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ (resp., $\mathcal{M}_{\mathbf{T}}^{(n)}(\Sigma)$ for all $n \geq 1$) as follows: $\mathcal{M}_{\mathbf{T}}^{(\wedge)}(\Sigma) = \{L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma) \ : \ (\forall L' \subseteq \Sigma^*) L \subset L' \Rightarrow L' \notin \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)\}$. (resp., $\mathcal{M}_{\mathbf{T}}^{(n)}(\Sigma) = \{L \in \mathcal{P}_{\mathbf{T}}^{(n)}(\Sigma) \ : \ (\forall L' \subseteq \Sigma^*) L \subset L' \Rightarrow L' \notin \mathcal{P}_{\mathbf{T}}^{(n)}(\Sigma)\}$.) In these definitions, $\subset$ indicates proper containment. By Zorn's Lemma, it is easy to see that for all $L \in \mathcal{P}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ (resp., $L \in \mathcal{P}_{\mathbf{T}}^{(n)}(\Sigma)$) there exists $L' \subseteq \mathcal{M}_{\mathbf{T}}^{(\wedge)}(\Sigma)$ (resp., $L' \in \mathcal{M}_{\mathbf{T}}^{(n)}(\Sigma)$) such that $L \subseteq L'$.

Our first result on maximality states that $\wedge$-equivalence carries over to maximal classes:

Lemma 20 *Let* $\mathbf{T}_1, \mathbf{T}_2 \subseteq 2^{\{0,1\}^*}$ *be hypersets of trajectories. Let* $n \geq 1$*. For all alphabets* Σ*,* $\mathbf{T}_1 \equiv_\wedge \mathbf{T}_2$ *(resp.,* $\mathbf{T}_1 \equiv_n \mathbf{T}_2$*), if and only if* $\mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) = \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$ *(resp.,* $\mathcal{M}_{\mathbf{T}_1}^{(n)}(\Sigma) = \mathcal{M}_{\mathbf{T}_1}^{(n)}(\Sigma)$*).*

PROOF. We establish the result only for $\wedge$-equivalence. The other case is left to the reader.

($\Rightarrow$): Assume that $\mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) \neq \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$. Without loss of generality, let $L \in \mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) - \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$. As $L \notin \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$, there are two cases: either $L \notin \mathcal{P}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$ or $L \in \mathcal{P}_{\mathbf{T}_2}^{(\wedge)}(\Sigma) - \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$.

In the first case, $L \notin \mathcal{P}_{\mathbf{T}_1}^{(\wedge)}(\Sigma)$ as $\mathbf{T}_1 \equiv_\wedge \mathbf{T}_2$. This contradicts that $L \in \mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) \subseteq \mathcal{P}_{\mathbf{T}_1}^{(\wedge)}(\Sigma)$. In the second case, by assumption, there exists $L' \in \mathcal{P}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$ such that $L \subset L'$. By assumption, $L' \in \mathcal{P}_{\mathbf{T}_1}^{(\wedge)}(\Sigma)$, again contradicting that $L \in \mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma)$.

($\Leftarrow$): Let $\mathcal{M}_{\mathbf{T}_1}^{(\wedge)}(\Sigma) = \mathcal{M}_{\mathbf{T}_2}^{(\wedge)}(\Sigma)$, and let $L \in \mathcal{P}_{\mathbf{T}_1}^{(\wedge)}(\Sigma)$. Then there exists $L' \in$

$\mathcal{M}^{(\wedge)}_{\mathbf{T}_1}(\Sigma) = \mathcal{M}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$ such that $L \subseteq L'$. Note that this implies that $L' \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$. But now the relation $L \subseteq L'$ implies that $L \in \mathcal{P}^{(\wedge)}_{\mathbf{T}_2}(\Sigma)$ as well, as required. $\square$

We note that it is decidable, given a regular language R and a regular set of trajectories T whether or not R is a maximal T-code [4]. This question is open for hypersets of regular sets of trajectories. Questions of maximality of code-like classes defined by the more general construct of *word operations* have also been investigated by Kari and Konstantinidis [17].

8 Conclusions

In this paper, we have begun examining classes of languages defined by hypersets of trajectories. This is an extension of those classes defined by a single set of trajectories, which themselves extend many classes of languages such as prefix codes, suffix codes, and hypercodes, among others.

Particular instances of codes defined by hypersets of trajectories have previously been studied in the literature, suggesting for example, the connection between the two classes of languages, $\mathcal{P}^{(\wedge)}_{\mathbf{T}}(\Sigma)$ and $\mathcal{P}^{(2)}_{\mathbf{T}}(\Sigma)$. We have extended these results. However, other surprising results have been obtained. In particular, we have shown that the regular languages membership problem for hypersets of trajectories is undecidable, even if all the sets of trajectories involved are regular.

Many open problems remain in this area. In particular, we have left open the decidability of equivalence for hypersets of regular sets of trajectories.

Acknowledgements

We thank the referees for their useful comments.

References

[1] Day, P.-H., and Shyr, H. Languages defined by some partial orders. *Soochow J. Math. 9* (1983), 53–62.

[2] Domaratzki, M. Deletion along trajectories. *Theor. Comp. Sci. 320*, 2–3 (2004), 293–313.

[3] DOMARATZKI, M. On the decidability of 2-infix-outfix codes. Tech. Rep. 2004-479, School of Computing, Queen's University, 2004.

[4] DOMARATZKI, M. Trajectory-based codes. *Acta Inf. 40*, 6–7 (2004), 491–527.

[5] DOMARATZKI, M. Trajectory-based embedding relations. *Fund. Inf. 59*, 4 (2004), 349–363.

[6] DOMARATZKI, M. Semantic shuffle on and deletion along trajectories. In *Developments in Language Theory* (2004), C. Calude, E. Calude, and M. Dineen, Eds., vol. 3340 of *LNCS*, Springer, pp. 163–174.

[7] DOMARATZKI, M. More words on trajectories. *Bull. Eur. Assoc. Theor. Comp. Sci. 86*, (2005), 107–145.

[8] DOMARATZKI, M., AND SALOMAA, K. Decidability of trajectory-based equations. *Theor. Comp. Sci. 345* (2005) 304–330.

[9] DOMARATZKI, M., AND SALOMAA, K. Codes defined by Multiple Trajectories. In *Automata and Formal Languages: 11th International Conference, AFL 2005* (2005), Z. Ésik and Z. Fülöp, eds., pp. 97–111.

[10] FAN, C.-M., SHYR, H., AND YU, S. S. d-Words and d-languages. *Acta Inf. 35* (1998), 709–727.

[11] ITO, M., JÜRGENSEN, H., SHYR, H., AND THIERRIN, G. n-prefix-suffix languages. *Intl. J. Comp. Math. 30* (1989), 37–56.

[12] JÜRGENSEN, H., AND KONSTANTINIDIS, S. Codes. pp. 511–600. In [28].

[13] JÜRGENSEN, H., SALOMAA, K., AND YU, S. Transducers and the decidability of independence in free monoids. *Theor. Comp. Sci. 134* (1994), 107–117.

[14] JÜRGENSEN, H., SHYR, H., AND THIERRIN, G. Codes and compatible partial orders on free monoids. In *Algebra and Order: Proceedings of the First International Symposium on Ordered Algebraic Structures, Luminy–Marseilles 1984* (1986), S. Wolfenstein, Ed., Heldermann Verlag, pp. 323–334.

[15] JÜRGENSEN, H., AND YU, S.-S. Relations on free monoids, their independent sets, and codes. *Int. J. Comput. Math. 40* (1991), 17–46.

[16] KADRIE, A., DARE, V., THOMAS, D., AND SUBRAMANIAN, K. Algebraic properties of the shuffle over ω-trajectories. *Inf. Proc. Letters 80*, 3 (2001), 139–144.

[17] KARI, L., AND KONSTANTINIDIS, S. Language equations, maximality and error-detection. *J. Comp. Sys. Sci. 70* (2005) 157–178.

[18] KARI, L., KONSTANTINIDIS, S., AND SOSÍK, P. On properties of bond-free DNA languages. *Theor. Comp. Sci. 334* (2005) 131–159.

[19] KARI, L., KONSTANTINIDIS, S., AND SOSÍK, P. Operations on trajectories with applications to coding and bioinformatics. *Int. J. Found. Comp. Sci. 16*, 3 (2005), 531–546.

[20] Kari, L., and Sosík, P. Aspects of shuffle and deletion on trajectories. *Theor. Comp. Sci. 332*, 1–3 (2005), 47–61.

[21] Long, D. *Study of Coding Theory and its Application to Cryptography.* PhD thesis, City University of Hong Kong, 2002.

[22] Long, D., Jia, W., Ma, J., and Zhou, D. k-p-infix codes and semaphore codes. *Disc. Appl. Math. 109* (2001), 237–252.

[23] Long, D., Ma, J., and Zhou, D. Structure of 3-infix-outfix maximal codes. *Theor. Comp. Sci. 188* (1997), 231–240.

[24] Mateescu, A. Splicing on routes: a framework of DNA computation. In *Unconventional Models of Computation* (1998), C. Calude, J. Casti, and M. Dinneen, Eds., Springer, pp. 273–285.

[25] Mateescu, A., and Mateescu, G. Associative and fair shuffle of ω-words. Tech. Rep. TUCS-TR-162, University of Turku, 1998.

[26] Mateescu, A., Rozenberg, G., and Salomaa, A. Shuffle on trajectories: Syntactic constraints. *Theor. Comp. Sci. 197* (1998), 1–56.

[27] Mateescu, A., Salomaa, K., and Yu, S. On fairness of many-dimensional trajectories. *J. Automata, Languages and Combinatorics 5* (2000), 145–157.

[28] Rozenberg, G., and Salomaa, A., Eds. *Handbook of Formal Languages, Vol. 1.* Springer-Verlag, 1997.

[29] Shyr, H. *Free Monoids and Languages.* Hon Min Book Company, Taichung, Taiwan, 2001.

[30] Shyr, H., and Thierrin, G. Codes and binary relations. In *Séminaire d'Algébre Paul Dubreil, Paris 1975–1976* (1977), A. Dold and B. Eckmann, Eds., vol. 586 of *Lecture Notes in Mathematics*, Springer-Verlag, pp. 180–188.

[31] Yu, S. S. d-Minimal languages. *Disc. Appl. Math. 89* (1998), 243–262.