Chapter 6

Trajectory-Based Codes

6.1 Introduction

The theory of codes is a fundamental area of formal language theory, with many important applications. The class of prefix codes is a particularly important subclass of codes, and is fundamentally linked to the nature of concatenation as the underlying operation. Further research in codes has considered the subclasses of codes which arise from replacing catenation with other, related operations, most notably shuffle (the hypercodes) and insertion (the outfix codes).

In this chapter, we generalize these results by considering T-codes. A T-code is any language L satisfying the equation $(L \sqcup\!\sqcup_T \Sigma^+) \cap L = \emptyset$. Thus, we consider the natural extension of prefix codes to all operations defined by shuffle on trajectories, and examine the properties of these classes of languages.

The idea of studying general classes of codes has received much attention in the literature (see, e.g., Shyr and Thierrin [186], Jürgensen *et al.* [99] and Jürgensen and Yu [100]). Further, the definition of a T-code which we present can also be formulated in dependency theoretic terms (see, e.g., Jürgensen and Konstantinidis [97] for a survey of dependency theory). Some of the results we have obtained can be proven by appealing to dependency theory, however, our proofs are simpler in our restricted situation.

In addition, there are works in the literature which consider the problem of defining codes based on arbitrary binary relations, see, e.g., the work of Jürgensen *et al.* [99] on codes defined by binary relations and Shyr and Thierrin [186] for work on so-called *strict* binary relations. We will see that we can also view T-codes as anti-chains under the natural binary relation defined by T.

With this research in mind, we nonetheless feel the framework of T-codes is useful in that it helps us to see results relating to codes defined by shuffle on trajectories in a new way. The restriction of considering only those codes defined by shuffle on trajectories gives us new insight into these classes, including prefix-, suffix-, bi(pre)fix-, infix-, outfix-, shuffle- and hyper-codes, by focusing our attention to classes of codes which are specific enough to allow reasoning on the associated sets of trajectories, but general enough to encompass all of the above interesting and well-studied classes of codes.

We also feel that introducing the idea of T-code will allow more unified results to be obtained on the various classes of codes, since specific conditions on sets of trajectories (i.e., languages) will be easier to obtain than more general conditions on arbitrary relations. In particular, we have obtained results which do not appear to have been considered before in the more general framework of dependency theory or binary relations.

Further, we note that the notion of T-codes is useful elsewhere in the study of iterated shuffle and deletion along trajectories, for instance, in analyzing the *shuffle-base* of certain languages. We examine this relationship in Section 8.9. Finally, the study of T-codes, much like the study of shuffle on trajectories in general, allows us to examine what assumptions must be made on an operation in order for certain results to follow. We find that even when these assumptions have been studied in the literature, the proofs obtained for the specific cases of shuffle on trajectories are often simpler.

We obtain several interesting results on T-codes. We generalize a result relating outfix and hyper-codes and the notion of (embedding-) convexity to all T-codes. Further, the known closure properties of shuffle on trajectories allow us to easily conclude positive decidability results for the problem of determining membership in classes of T-codes (including maximal T-codes), which were previously determined by ad-hoc constructions in the literature.

We note that recently, a more general concept than T-codes has been independently introduced by Kari *et al.* [108], motivated by the bonding of strands of DNA and DNA computing. Their framework, called *bond-free properties*, is also a general setting which involves shuffle on trajectories. Generally, the motivations for our work and those of the work by Kari *et al.* are different, and the decidability results which are similar are noted below.

6.2 Definitions

Recall that a non-empty language L is a *code* if $u_1 u_2 \cdots u_m = v_1 v_2 \cdots v_n$ where $u_i, v_j \in L$ for $1 \le i \le m$ and $1 \le j \le n$ implies that $n = m$ and $u_i = v_i$ for $1 \le i \le n$. For background on codes, we refer the reader to Berstel and Perrin [18], Jürgensen and Konstantinidis [97] or Shyr [184].

We now come to the main definition of this chapter. Let $L \subseteq \Sigma^+$ be a language. Then, for any $T \subseteq \{0,1\}^*$, we say that L is a *T-code* if L is non-empty and $(L \sqcup\!\sqcup_T \Sigma^+) \cap L = \emptyset$. If Σ is an alphabet and $T \subseteq \{0,1\}^*$, let $\mathcal{P}_T(\Sigma)$ denote the set of all T-codes over Σ. If Σ is understood, we will denote the set of T-codes over Σ by $\mathcal{P}_T$.

There has been much research into the idea of T-codes for particular $T \subseteq \{0,1\}^*$, including

(a) prefix codes, corresponding to $T = 0^*1^*$ (concatenation);

(b) suffix codes, corresponding to $T = 1^*0^*$ (anti-catenation);

(c) biprefix (or bifix) codes, corresponding to $T = 0^*1^* + 0^*1^*$ (bi-catenation);

(d) outfix and infix codes, corresponding to $T = 0^*1^*0^*$ (insertion) and $T = 1^*0^*1^*$, (bi-polar insertion) respectively;

(e) shuffle-codes, corresponding to bounded sets of trajectories such as

(e-i) $T = (0^*1^*)^n$ for fixed $n \ge 1$ (prefix codes of index n);

(e-ii) $T = (1^*0^*)^n$ for fixed $n \ge 1$ (suffix codes of index n);

(e-iii) $T = 1^*(0^*1^*)^n$ for fixed $n \ge 1$ (infix codes of index n);

(e-iv) $T = (0^*1^*)^n0^*$ for fixed $n \ge 1$ (outfix codes of index n);

(f) hypercodes, corresponding to $T = (0+1)^*$ (arbitrary shuffle);

(g) k-codes, corresponding to $T = 0^*1^*0^{\leq k}$ (k-catenation, see Kari and Thierrin [114]) for fixed $k \geq 0$; and

(h) for arbitrary $k \geq 1$, codes defined by the sets of trajectories $PP_k = 0^* + (0^*1^*)^{k-1}0^*1^+$, $PS_k = 0^* + 1^+0^*(1^*0^*)^{k-1}$, $PI_k = 0^* + (1^*0^*)^k 1^+$, $SI_k = 0^* + 1^+(0^*1^*)^k$, $PB_k = PP_k \cup PS_k$ and $BI_k = PI_k \cup SI_k$, see Long [135], or Ito *et al.* [77] for PI_1, SI_1.

For a list of references related to (a)–(f), see Jürgensen and Konstantinidis [97, pp. 549–553]. In this chapter, we let

$$H = (0+1)^*, \tag{6.1}$$

$$P = 0^*1^*, \tag{6.2}$$

$$S = 1^*0^*, \tag{6.3}$$

$$I = 1^*0^*1^*, \tag{6.4}$$

$$O = 0^*1^*0^*, \text{ and} \tag{6.5}$$

$$B = P \cup S. \tag{6.6}$$

6.3 General Properties of T-codes

We can give two alternate characterizations of T-codes in terms of the left and right inverses of shuffle on trajectories. These are given via the morphisms $\tau, \pi : \{0,1\}^* \to \{i,d\}^*$ defined by $\tau(0) = i$, $\tau(1) = d$, $\pi(0) = d$ and $\pi(1) = i$. We can easily prove the following two equalities by appealing to Theorems 5.8.1 and 5.8.2. In particular, we have for all $T \subseteq \{0,1\}^*$, and all Σ,

$$\mathcal{P}_T(\Sigma) = \{L \subseteq \Sigma^+ : (L \rightsquigarrow_{\tau(T)} \Sigma^+) \cap L = \emptyset\}, \tag{6.7}$$

$$\mathcal{P}_T(\Sigma) = \{L \subseteq \Sigma^+ : L \rightsquigarrow_{\pi(T)} L \subseteq \{\epsilon\}\}. \tag{6.8}$$

For some particular T, these characterizations are well-known, e.g., (6.7) for $T = 0^*1^*$ is given by Berstel and Perrin [18, Prop. II.1.1.(ii)].

We now note that the term T-code is somewhat of a misnomer: some T-codes are not codes. However, we feel that as T-codes are the natural analogues of prefix codes when catenation is replaced by $\sqcup\!\sqcup_T$, the term T-code is appropriate. The following example shows how T-codes can fail to be codes:

Example 6.3.1: Let $T = (01)^*$. Then $\sqcup\!\sqcup_T$ corresponds to perfect shuffle (also known as balanced literal shuffle). Then note that $L = \{aa, bb, aabb\}$ is a T-code: there is no way to perfectly shuffle aa (resp., bb) and any other word of length 2 to get $aabb$. However, L is not a code: $aa \cdot bb = aabb$. □

The following states that more restrictive sets of trajectories (potentially) result in more languages being T-codes; the proof is immediate:

Lemma 6.3.2 *Let $T_1 \subseteq T_2 \subseteq \{0, 1\}^*$. Then for all Σ, $\mathcal{P}_{T_1}(\Sigma) \supseteq \mathcal{P}_{T_2}(\Sigma)$.*

By the fact that all prefix codes are codes, we conclude the following, which complements Example 6.3.1:

Corollary 6.3.3 *Let $T \supseteq 0^*1^*$. Then every T-code is a code.*

Let $\mathcal{P}_{\text{CODE}}$ denote the set of all codes. We now show that for all $T \subseteq \{0, 1\}^*$, $\mathcal{P}_T \neq \mathcal{P}_{\text{CODE}}$. We will require the following well-known characterization of two element codes (see, e.g., Berstel and Perrin [18, Cor. 2.9]):

Theorem 6.3.4 *Let $L = \{x_1, x_2\} \subseteq \Sigma^+$. Then L is not a code if and only if there exist $z \in \Sigma^+$, $i, j \in \mathbb{N}^+$ such that $x_1 = z^i$ and $x_2 = z^j$.*

Lemma 6.3.5 *Let $T \subseteq \{0, 1\}^*$. Then $\mathcal{P}_T(\Sigma) \neq \mathcal{P}_{\text{CODE}}(\Sigma)$ for all Σ with $|\Sigma| > 1$.*

Proof. Let $T \subseteq \{0, 1\}^*$. If $T \subseteq 0^* + 1^*$, then $\mathcal{P}_T = \mathcal{P}_\emptyset = 2^{\Sigma^+} - \{\emptyset\}$ (the first equality will become clear after Theorem 6.3.7 below), which is clearly not the set of codes.

Thus, we can assume that there is some $t \in T$ with $|t|_1, |t|_0 > 0$. Let $n = |t|_0$. Consider that $t \in 0^n \sqcup\!\sqcup_t \{0,1\}^+$. Thus $L = \{t, 0^n\} \subseteq \{0,1\}^+$ is not a T-code.

If L is not a code, then t and 0^n are powers of the same word, i.e., $t \in 0^*$. This contradicts our choice of t. Thus, L is a code. ■

We also observe that $\mathcal{P}_{T_1} \cap \mathcal{P}_{T_2} = \mathcal{P}_{T_1 \cup T_2}$. We note that the dual case does not hold. In the case of $\mathcal{P}_{T_1 \cap T_2}$, we have the inclusion $\mathcal{P}_{T_1} \cap \mathcal{P}_{T_2} \subseteq \mathcal{P}_{T_1 \cap T_2}$. But of course equality does not hold in general. For example, with $T_1 = 0^*1^*$ and $T_2 = 1^*0^*$, $\mathcal{P}_{T_1 \cap T_2} = \mathcal{P}_{0^*+1^*} = \mathcal{P}_{\emptyset} = 2^{\Sigma^+} - \{\emptyset\}$ (the second equality will be established in Theorem 6.3.7 below). However, $\mathcal{P}_{T_1} \cap \mathcal{P}_{T_2} = \mathcal{P}_{T_1 \cup T_2}$, the set of biprefix codes.

We can also ask if $T_1 \subset T_2$ ($\subset$ denotes proper inclusion) implies that $\mathcal{P}_{T_1} \supset \mathcal{P}_{T_2}$. The answer is yes, as long as the difference between T_2 and T_1 contains non-unary words.

Theorem 6.3.6 *Let $T_1 \subset T_2$ be such that $(T_2 - T_1) \cap \overline{0^* + 1^*} \neq \emptyset$. Then for all Σ with $|\Sigma| \geq 2$, $\mathcal{P}_{T_1}(\Sigma) \supset \mathcal{P}_{T_2}(\Sigma)$.*

Proof. Let $t \in (T_2 - T_1) \cap \overline{0^* + 1^*}$. Let t_0, t_1 be defined by $t_0 = 0^{|t|_0}$ and $t_1 = 1^{|t|_1}$. Then note that $t_0, t_1 \neq \epsilon$, by our choice of t. Thus, we have that $\{t, t_0\} \subseteq \{0,1\}^+$. We claim that $L_t = \{t, t_0\} \in \mathcal{P}_{T_1} - \mathcal{P}_{T_2}$.

To see that $L_t \notin \mathcal{P}_{T_2}$, note that $t \in t_0 \sqcup\!\sqcup_t t_1$. As $t \in T_2$ and $t_1 \neq \epsilon$, L_t is not a T_2-code. Assume that L_t is not a T_1-code. As $|t| > |t_0|$, the only way that L_t can fail to be a T_1-code is if there exists $x \in \{0,1\}^+$ such that $t \in t_0 \sqcup\!\sqcup_{T_1} x$. By definition, as $|t|_0 = |t_0|$, we must have that $x = t_1 = 1^{|t|_1}$. But $t \in t_0 \sqcup\!\sqcup_{T_1} t_1$ only if $t \in T_1$, which is not the case. ■

Theorem 6.3.7 *Let $T_1 \subset T_2$ and $T_2 - T_1 \subseteq 1^* + 0^*$. Then for all Σ with $|\Sigma| > 1$, $\mathcal{P}_{T_1}(\Sigma) = \mathcal{P}_{T_2}(\Sigma)$.*

Proof. Assume, contrary to what we want to prove, that $L \subseteq \Sigma^+$ is a T_1-code which is not a T_2-code. As L is not a T_2-code, there exist $x, z \in L$, $y \in \Sigma^+$ and $t \in T_2$ such that $z \in x \sqcup\!\sqcup_t y$. As L is a T_1-code, $z \notin x \sqcup\!\sqcup_{T_1} y$. Thus $t \notin T_1$. By assumption, this implies that $t \in 1^* + 0^*$.

If $t \in 1^*$, then by definition of $\sqcup\!\sqcup_T$, $z \in x \sqcup\!\sqcup_t y$ implies that $x = \epsilon$, contrary to our choice of L. If $t \in 0^*$, then by definition, $y = \epsilon$, contrary to our choice of y. In either case, we have arrived at a contradiction. ■

Thus, we have completely characterized when reducing a set of trajectories corresponds to an increase in the languages which are T-codes. In particular, we note the following corollary:

Corollary 6.3.8 *Let $T_1, T_2 \subseteq \{0, 1\}^*$ be regular sets of trajectories. Then it is decidable whether $\mathcal{P}_{T_1} = \mathcal{P}_{T_2}$.*

Proof. We note that $\mathcal{P}_{T_1} = \mathcal{P}_{T_2}$ if and only if $(T_1 - T_2) \cup (T_2 - T_1) \subseteq 0^* + 1^*$. Since T_1, T_2 are regular, so is $(T_1 - T_2) \cup (T_2 - T_1)$, and the inclusion is decidable. ■

We now examine further questions of decidability.

Lemma 6.3.9 *Let $T \subseteq \{0, 1\}^*$ be a fixed CF set of trajectories. Then given a regular language L, it is decidable whether L is a T-code.*

Proof. Since L is regular and T is a CFL, $L \sqcup\!\sqcup_T \Sigma^+$, and $(L \sqcup\!\sqcup_T \Sigma^+) \cap L$ are CFLs. Thus, we can test whether $(L \sqcup\!\sqcup_T \Sigma^+) \cap L = \emptyset$, which precisely defines L being a T-code. ■

This result can also be proved using dependency theory. As every $T \subseteq \{0, 1\}^*$ defines a 3-dependence system, and every context-free T defines a dependence system whose associated support can be accepted by a 3-tape PDA, the problem of determining membership in $\mathcal{P}_T$ is decidable; see Jürgensen and Konstantinidis [97, Sect. 9] for details. Further, Kari *et al.* [108, Thm. 4.7] establish a similar decidability result in their framework of *bond-free properties*. When translated to our setting, it states that given T, R regular, we can decide if $R \in \mathcal{P}_T$.

A class of languages $\mathcal{C}$ is said to have *decidable membership problem* if, given $L \subseteq \Sigma^*$ with $L \in \mathcal{C}$, it is decidable whether $x \in L$ for an arbitrary $x \in \Sigma^*$. We have the following positive decidability result:

Lemma 6.3.10 *Let $\mathcal{C}$ be a class of languages with decidable membership. Let $T \subseteq \{0, 1\}^*$ be a set of trajectories such that $T \in \mathcal{C}$. Then given a finite language F, it is decidable whether $F \in \mathcal{P}_T$.*

Proof. Let $F \subseteq \Sigma^+$ be a finite set. Let $n = \max\{|x| : x \in F\}$. Since membership in T is decidable, we can test all $t \in \{0, 1\}^{\leq n}$ for membership in T. Thus, we can effectively compute $T^{\leq n} = T \cap \{0, 1\}^{\leq n}$. It is easily observed that $F \cap (F \sqcup\!\sqcup_{T^{\leq n}} L) = F \cap (F \sqcup\!\sqcup_T L)$ for all L.

Since $F, T^{\leq n}, \Sigma^+$ are regular, we can test $F \cap (F \sqcup\!\sqcup_{T^{\leq n}} \Sigma^+) = \emptyset$. Thus, the result follows. ■

We conclude with the following method of constructing a T-code from an arbitrary language.

Lemma 6.3.11 *Let $T \subseteq \{0, 1\}^*$. Let $L \subseteq \Sigma^+$ be a non-empty language. Then $L_0 = L - (L \sqcup\!\sqcup_T \Sigma^+) \in \mathcal{P}_T(\Sigma)$.*

Proof. As $L_0 \subseteq L$ and $\sqcup\!\sqcup_T$ is a monotone operation, $(L_0 \sqcup\!\sqcup_T \Sigma^+) \subseteq (L \sqcup\!\sqcup_T \Sigma^+)$. Thus, $L_0 \cap (L_0 \sqcup\!\sqcup_T \Sigma^+) \subseteq L_0 \cap (L \sqcup\!\sqcup_T \Sigma^+)$ and $L_0 \cap (L \sqcup\!\sqcup_T \Sigma^+) = \emptyset$ by definition of L_0. ■

The following is proven in exactly the same manner as Lemma 6.3.11:

Lemma 6.3.12 *Let $T \subseteq \{0, 1\}^*$. Let $L \subseteq \Sigma^+$ be a non-empty language. Then $L_0 = L - (L \rightsquigarrow_{\tau(T)} \Sigma^+) \in \mathcal{P}_T(\Sigma)$.*

6.4 The Binary Relation defined by Trajectories

We can also define T-codes by appealing to a definition based on binary relations. In particular, for $T \subseteq \{0, 1\}^*$, define ω_T as follows: for all $x, y \in \Sigma^*$,

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

Then it is clear that $L \subseteq \Sigma^+$ is a T-code if and only if L is an anti-chain under ω_T (i.e, $x, y \in L$ and $x \; \omega_T \; y$ implies $x = y$).

We note that the relation analogous to ω_T for infinite words and ω-trajectories was defined by Kadrie *et al.* [101], and its properties were briefly investigated. Kadrie *et al.* do not investigate the analogous relation with the same amount of detail as below and do not appear to be motivated by the theory of codes.

We immediately note that if $T_1, T_2 \subseteq \{0,1\}^*$ are sets of trajectories, there is not necessarily a set of trajectories T such that $\omega_T = \omega_{T_1} \cap \omega_{T_2}$, i.e., such that $x \ \omega_T \ y \iff (x \ \omega_{T_1} \ y) \wedge (x \ \omega_{T_2} \ y)$. For instance, for $P = 0^*1^*$ and $S = 1^*0^*$, the relation $\omega_P \cap \omega_S$ is given by $\leq_d$, where $x \leq_d y$ if and only if there exist $u, v \in \Sigma^*$ such that $y = xu = vx$. This relation cannot be represented by a set of trajectories:

Lemma 6.4.1 *For all $T \subseteq \{0,1\}^*$, $\omega_T \neq \leq_d$.*

Proof. Assume that there exists $T \subseteq \{0,1\}^*$ such that $\omega_T = \leq_d$. Consider $L_0 = \{0, 00\}$. As $0 \leq_d 00$, we must have that $0 \ \omega_T \ 00$. Thus, $00 \in 0 \sqcup\!\sqcup_T 0$ and $\{01, 10\} \cap T \neq \emptyset$. Thus, without loss of generality assume that $01 \in T$. The case $10 \in T$ is similar.

Consider now $L_1 = \{0, 01\}$. We observe that L_1 is an anti-chain under $\leq_d$, i.e., $0 \leq_d 01$ does not hold. However, $01 \in 0 \sqcup\!\sqcup_T 1$. Thus, $0 \ \omega_T \ 01$, and $\omega_T \neq \leq_d$. ■

For a discussion of $\leq_d$, see Shyr [184, Ch. 8]. We now recall some of the properties of the binary relations ω_T that will be useful. In what follows, we will refer to T having a property P if and only if ω_T has property P.

6.4.1 Anti-symmetry

Recall that a binary relation ρ is *anti-symmetric* if $x \ \rho \ y$ and $y \ \rho \ x$ implies $x = y$. We note that ω_T always gives an anti-symmetric binary relation:

Lemma 6.4.2 *Let $T \subseteq \{0,1\}^*$. The relation ω_T is anti-symmetric.*

Proof. Let $x, y \in \Sigma^*$ be such that $x \ \omega_T \ y \ \omega_T \ x$. Then let $t_1, t_2 \in T$ and $\alpha, \beta \in \Sigma^*$ be such that $x \in y \sqcup\!\sqcup_{t_1} \alpha$ and $y \in x \sqcup\!\sqcup_{t_2} \beta$. By definition of shuffle on trajectories, $|x| = |y| + |\alpha|$ and $|y| = |x| + |\beta|$. Thus, $|\alpha| = |\beta| = 0$, i.e., $\alpha = \beta = \epsilon$. But now $x \in y \sqcup\!\sqcup_{t_1} \epsilon$, which implies that $x = y$, again by definition of shuffle on trajectories. ■

6.4.2 Reflexivity

Recall that a binary relation ρ on Σ^* is *reflexive* if $x \; \rho \; x$ for all $x \in \Sigma^*$.

Lemma 6.4.3 *Let $T \subseteq \{0, 1\}^*$. Then T is reflexive if and only if $0^* \subseteq T$.*

Proof. Let $0^* \subseteq T$. Then $x \in x \sqcup\!\sqcup_{0^{|x|}} \epsilon$, i.e., $x \; \omega_T \; x$. Thus ω_T is reflexive. For the converse, let $x \in x \sqcup\!\sqcup_T \Sigma^*$ for all $x \in \Sigma^*$. Then clearly $0^{|x|} \in T$ for all $x \in \Sigma^*$, which implies $0^* \subseteq T$. ■

Corollary 6.4.4 *Given a CF set $T \subseteq \{0, 1\}^*$ of trajectories, it is decidable whether T is reflexive.*

Proof. Let $T' = T \cap 0^*$, which is a unary CFL, and thus regular. In fact, if T is effectively context-free, then T' is effectively regular. We can then test the equality $0^* = T'$. ■

6.4.3 Positivity

A binary relation ρ on Σ^* is said to be *positive* if $\epsilon \; \rho \; x$ for all $x \in \Sigma^*$.

Lemma 6.4.5 *Let $T \subseteq \{0, 1\}^*$. Then T is positive if and only if $1^* \subseteq T$.*

Proof. Let $1^* \subseteq T$. Then $u \in \epsilon \sqcup\!\sqcup_{1^{|u|}} u$ for all $u \in \Sigma^*$, whereby $\epsilon \; \omega_T \; u$, as $1^{|u|} \in T$. The reverse implication is similarly established. ■

Corollary 6.4.6 *Given a CF set $T \subseteq \{0, 1\}^*$ of trajectories, it is decidable whether T is positive.*

6.4.4 ST-Strictness

Shyr and Thierrin [186] define the concept of a *strict* binary relation. To avoid confusion with the concept of a *strict ordering* (see, e.g., Choffrut and Karhumäki [24, Sect. 7.1]), we will call a binary relation ρ on Σ^* *ST-strict* if it satisfies the following four properties:

(a) ρ is reflexive;

(b) ρ is positive;

(c) for all $u, v \in \Sigma^*$, $u \,\rho\, v$ implies $|u| \leq |v|$;

(d) for all $u, v \in \Sigma^*$, $u \,\rho\, v$ and $|u| = |v|$ implies $u = v$.

We now consider T such that ω_T is ST-strict. We first note that conditions (c) and (d) are satisfied by all T. Indeed, if $u \,\omega_T\, v$, then $v \in u \sqcup\!\sqcup_T \Sigma^*$, which implies that $|v| \geq |u|$. Further, if $|u| = |v|$, then $u \,\omega_T\, v$ implies that $v \in u \sqcup\!\sqcup_T \epsilon$, which implies that $u = v$.

Thus, as we already have necessary and sufficient conditions on T being reflexive and positive, the following results are immediate:

Corollary 6.4.7 *Let $T \subseteq \{0,1\}^*$. Then T is ST-strict if and only if $0^* + 1^* \subseteq T$.*

Corollary 6.4.8 *Given a CF set $T \subseteq \{0,1\}^*$ of trajectories, it is decidable whether T is ST-strict.*

Corollary 6.4.9 *Let $T_1, T_2 \subseteq \{0,1\}^*$ be ST-strict. Then $\mathcal{P}_{T_1} = \mathcal{P}_{T_2}$ if and only if $T_1 = T_2$.*

6.4.5 Cancellativity

A binary relation ρ on Σ^* is said to be *left-cancellative* (resp., *right-cancellative*) if $uv \,\rho\, ux$ implies $v \,\rho\, x$ (resp., $vu \,\rho\, xu$ implies $v \,\rho\, x$) for all $u, v, x \in \Sigma^*$. The relation ρ is *cancellative* if it is both left- and right-cancellative.

Given $T \subseteq \{0,1\}^*$, we define two sets of trajectories, $s(T), p(T) \subseteq \{0,1\}^*$, as follows:

$$\begin{aligned} p(T) &= \{t_1 1^j \;:\; t_1 t_2 \in T, 0 \leq j \leq |t_2|\}, \\ s(T) &= \{1^j t_2 \;:\; t_1 t_2 \in T, 0 \leq j \leq |t_1|\}. \end{aligned}$$

Lemma 6.4.10 *Let $T \subseteq \{0,1\}^*$. Then T is left-cancellative (resp., right-cancellative) if $s(T) \subseteq T$ (resp., $p(T) \subseteq T$).*

Proof. We establish the result for left-cancellativity only; the other case is symmetric. Let $s(T) \subseteq T$. Then let $u, v, x \in \Sigma^*$ be such that $uv \,\omega_T\, ux$. Let $t \in T$ and $\alpha \in \Sigma^*$ chosen so that $ux \in uv \sqcup\!\sqcup_t \alpha$. Write $t = t_1 t_2$ and $\alpha = \alpha_1 \alpha_2$ so that $ux \in (u \sqcup\!\sqcup_{t_1} \alpha_1)(v \sqcup\!\sqcup_{t_2} \alpha_2)$. Let $\beta_1, \beta_2 \in \Sigma^*$ be chosen so that

$\beta_1 \in u ⧢_{t_1} \alpha_1$, $\beta_2 \in v ⧢_{t_2} \alpha_2$ and $ux = \beta_1\beta_2$. As $|\beta_1| = |u| + |\alpha_1| \geq |u|$, there exists $\gamma \in \Sigma^*$ such that $u\gamma = \beta_1$ and $x = \gamma\beta_2$. Note that $|\gamma| \leq |\beta_1| = |t_1|$. Thus,

$$x \in \gamma (v ⧢_{t_2} \alpha_2).$$

Let $t_3 = 1^{|\gamma|}t_2 \in s(T)$. By assumption, $t_3 \in T$. Further,

$$x \in v ⧢_{t_3} \gamma\alpha_2.$$

We conclude that $v \; \omega_T \; x$. ■

Corollary 6.4.11 *Let $T \subseteq \{0, 1\}^*$. If $s(T) \cup p(T) \subseteq T$, then T is cancellative.*

We now consider a condition of Jürgensen *et al.* [99]. Say that a binary relation ρ on Σ^* is *leviesque* if $uv \; \rho \; xy$ implies that $u \; \rho \; x$ or $v \; \rho \; y$, for all $u, v, x, y \in \Sigma^*$.

Lemma 6.4.12 *Let $T \subseteq \{0, 1\}^*$. If $s(T) \cup p(T) \subseteq T$, then T is leviesque.*

Proof. Let $rs \; \omega_T \; xy$. Then there exist $t \in T$ and $\alpha \in \Sigma^*$ such that $xy \in rs ⧢_t \alpha$. Then there exist factorizations $t = t_1t_2$, $\alpha = \alpha_1\alpha_2$ such that $xy \in (r ⧢_{t_1} \alpha_1)(s ⧢_{t_2} \alpha_2)$. Let $\beta_1, \beta_2 \in \Sigma^*$ be such that $\beta_1 \in r ⧢_{t_1} \alpha_1$, $\beta_2 \in s ⧢_{t_2} \alpha_2$ and $xy = \beta_1\beta_2$. There are two cases:

(i) If $|x| \geq |\beta_1|$, then there exists $\gamma \in \Sigma^*$ such that $x = \beta_1\gamma$ and $\gamma y = \beta_2$. Note that $|\gamma| \leq |\beta_2| = |t_2|$. Consider that $x = \beta_1\gamma \in (r ⧢_{t_1 1^{|\gamma|}} \alpha_1\gamma)$. As $t_1 1^{|\gamma|} \in p(T) \subseteq T$, $r \; \omega_T \; x$.

(ii) If $|x| \leq |\beta_1|$, there exists $\gamma \in \Sigma^*$ such that $x\gamma = \beta_1$ and $y = \gamma\beta_2$. Note that $|\gamma| \leq |\beta_1| = |t_1|$. In this case, $y = \gamma\beta_2 \in (s ⧢_{1^{|\gamma|}t_2} \gamma\alpha_2)$. Thus, as $1^{|\gamma|}t_2 \in s(T) \subseteq T$, we have $s \; \omega_T \; y$.

Thus, $rs \; \omega_T \; xy$ implies $(r \; \omega_T \; x)$ or $(s \; \omega_T \; y)$. ■

6.4.6 Compatibility

Let ρ be a binary relation on Σ^*. Then we say that ρ is *left-compatible* (resp., *right-compatible*) if, for all $u, v, w \in \Sigma^*$, $u \; \rho \; v$ implies that $wu \; \rho \; wv$ (resp., $uw \; \rho \; vw$). If ρ is both left- and right-compatible, we say it is *compatible*.

Lemma 6.4.13 *Let $T \subseteq \{0,1\}^*$. Then T is right-compatible (resp., left-compatible) if and only if $T0^* \subseteq T$ (resp., $0^*T \subseteq T$).*

Proof. We establish the result for right-compatibility. The result for left-compatibility is symmetrical.

Let $T0^* \subseteq T$. Let $u, v, w \in \Sigma^*$ with $u \; \omega_T \; v$. Then there exist $t \in T$ and $\alpha \in \Sigma^*$ such that $v \in u \sqcup\!\sqcup_t \alpha$. As $t' = t0^{|w|} \in T$, $vw \in uw \sqcup\!\sqcup_{t'} \alpha$. Thus $uw \; \omega_T \; vw$.

Assume that $T0^*$ is not a subset of T. Then there exist $t \in T$ and $i \in \mathbb{N}$ such that $t0^i \notin T$. Let $j = |t|_0$ and $k = |t|_1$. Consider that $0^j \; \omega_T \; t$, as $t \in 0^j \sqcup\!\sqcup_t 1^k$. However, $0^j \cdot 0^i \; \omega_T \; t \cdot 0^i$ does not hold, as $t0^i \in 0^{j+i} \sqcup\!\sqcup_T 1^k$ would imply that $t0^i \in T$. Thus, T is not right-compatible. ■

The following corollary is immediate; it is identical to the condition that $T0^* \cup 0^*T \subseteq T$:

Corollary 6.4.14 *Let $T \subseteq \{0,1\}^*$. Then T is compatible if and only if $0^*T0^* \subseteq T$.*

Corollary 6.4.15 *Given a regular set $T \subseteq \{0,1\}^*$ of trajectories, it is decidable whether T is (left- or right-) compatible.*

Lemma 6.4.16 *Given an LCF set $T \subseteq \{0,1\}^*$ of trajectories, it is undecidable whether T is (left- or right-) compatible.*

Proof. We prove only the case for left-compatibility; the other cases are similar and are left to the reader. We apply a meta-theorem of Hunt and Rosenkrantz (Theorem 2.5.3). First, we note that $T = \{0,1\}^*$ is left-compatible.

Let $T = \{0^n 1^n \; : \; n \geq 0\}$. We claim that there is no LCF set $T' \subseteq \{0,1\}^*$ of trajectories and trajectory $t \in \{0,1\}^+$ such that $T = T'/t$. Assume that there were such T', t. Then as $\epsilon \in T = T'/t$, we must have $t \in T'$. As T' is left-compatible, we have that $0t \in T'$. Thus $0 \in T'/t = T$, a contradiction. Thus, the set

$$\{T \; : \; \exists \text{ left-compatible LCF } T' \subseteq \{0,1\}^*, t \in \{0,1\}^+ \text{ such that } T = T'/t\}$$

is a proper subset of the LCFLs. Therefore, we may apply Theorem 2.5.3, and it is undecidable whether a given LCF set of trajectories is left-compatible. ■

Recall the definitions of P, S and O given by (6.2), (6.3) and (6.5). Let $\mathcal{P}_P, \mathcal{P}_S, \mathcal{P}_O$ be the class of prefix, suffix and outfix codes. We can conclude the following corollary about positive T which satisfy compatibility conditions. Parts (a) and (b) of the following result have been established for all partial orders by Jürgensen *et al.* [99]; the proofs are immediate in our case:

Corollary 6.4.17 *Let $T \subseteq \{0,1\}^*$ be positive. Then the following hold:*

(a) if T is left-compatible, then $\mathcal{P}_T \subseteq \mathcal{P}_P$;

(b) if T is right-compatible, then $\mathcal{P}_T \subseteq \mathcal{P}_S$;

(c) if T is compatible, then $\mathcal{P}_T \subseteq \mathcal{P}_O$.

Furthermore, in each case equality of the classes holds if and only if it holds for the sets of trajectories involved.

Proof. We prove (b); the rest are similar. If T is positive then $1^* \subseteq T$. If T is right compatible, then $T0^* \subseteq T$. Thus, $S = 1^*0^* \subseteq T$. The inclusions thus hold by Lemma 6.3.2; for the equalities, we note that P, S, O are ST-strict and for each of (a),(b) and (c), T is also ST-strict. ■

6.4.7 Transitivity

Recall that a binary relation ρ on Σ^* is said to be *transitive* if $x \,\rho\, y$ and $y \,\rho\, z$ imply that $x \,\rho\, z$ for all $x, y, z \in \Sigma^*$. We now consider conditions on T which will ensure that ω_T is a transitive relation. Transitivity is often, but not always, a property of the binary relations defining the classic code classes. For instance, both bi-prefix and outfix codes are defined by binary relations which are not transitive, and hence not a partial order. We now give necessary and sufficient conditions on a set T of trajectories defining a transitive binary relation.

First, we define three morphisms we will need. Let $D = \{x, y, z\}$ and $\varphi, \sigma, \psi : D^* \to \{0,1\}^*$

be the morphisms given by

$$\begin{array}{rclrclrcl} \varphi(x) & = & 0, & \sigma(x) & = & 0, & \psi(x) & = & 0, \\ \varphi(y) & = & 0, & \sigma(y) & = & 1, & \psi(y) & = & 1, \\ \varphi(z) & = & 1, & \sigma(z) & = & \epsilon, & \psi(z) & = & 1. \end{array}$$

Note that these morphisms are similar to the substitutions defined by Mateescu *et al.* [147], whose purpose is to give necessary and sufficient conditions on a set T of trajectories defining an associative operation. Indeed, our condition is a weakening of their conditions, which, intuitively, reflects the fact that any associative operation $\sqcup\!\sqcup_T$ defines a transitive binary relation ω_T (note, however, that $T = 1^*0^*1^*$ is transitive but not associative).

Theorem 6.4.18 *Let $T \subseteq \{0,1\}^*$. Then T is transitive if and only if*

$$\psi(\varphi^{-1}(T) \cap \sigma^{-1}(T)) \subseteq T. \tag{6.9}$$

Proof. ($\Leftarrow$): Let T define a transitive binary relation. Let $w \in \psi(\varphi^{-1}(T) \cap \sigma^{-1}(T))$. Then there exist $t_1, t_2 \in T$ such that $w \in \psi(\varphi^{-1}(t_1) \cap \sigma^{-1}(t_2))$. Let $t \in \varphi^{-1}(t_1) \cap \sigma^{-1}(t_2)$ be chosen so that $w \in \psi(t)$.

Consider t_1. Let $n \in \mathbb{N}$ and $\alpha_i, \beta_i \in \mathbb{N}$ be chosen for $1 \le i \le n$ so that

$$t_1 = \prod_{i=1}^{n} 0^{\alpha_i} 1^{\beta_i}.$$

Note that

$$\varphi^{-1}(t_1) = \prod_{i=1}^{n} (x+y)^{\alpha_i} z^{\beta_i}.$$

As $t \in \varphi^{-1}(t_1)$ and $t \in \sigma^{-1}(t_2)$, by definition of σ, we must have that $t_2 = \prod_{i=1}^{n} s_i$ for $s_i \in \{0,1\}^*$ satisfying $|s_i| = \alpha_i$. Thus, we have that $|t_2| = |t_1|_0$. Furthermore, $t \in (x^{p_1} \sqcup\!\sqcup_{t_2} y^{p_2}) \sqcup\!\sqcup_{t_1} z^{p_3}$, where $p_1 = |t_2|_0$, $p_2 = |t_2|_1$ and $p_3 = |t_1|_1$. Consider now that $w \in \psi(t)$, so that

$$w \in (0^{p_1} \sqcup\!\sqcup_{t_2} 1^{p_2}) \sqcup\!\sqcup_{t_1} 1^{p_3}.$$

Clearly, $0^{p_1} \sqcup\!\sqcup_{t_2} 1^{p_2} = t_2$. Thus, $w \in t_2 \sqcup\!\sqcup_{t_1} 1^{p_3}$, as well. By definition, we then have that $0^{p_1}\ \omega_T\ t_2\ \omega_T\ w$. By the transitivity of T, $0^{p_1}\ \omega_T\ w$, i.e.,

$$w \in 0^{p_1} \sqcup\!\sqcup_T \{0,1\}^*.$$

Note that $|w|_1 = p_2 + p_3$ and $|w|_0 = p_1$. The only word v over $\{0,1\}$ such that $w \in 0^{p_1} \sqcup\!\sqcup_T v$ is $v = 1^{p_2+p_3}$ (regardless of T). That is, $w \in 0^{p_1} \sqcup\!\sqcup_T 1^{p_2+p_3}$. But from this, we must have that $w \in T$. Thus, we have that $\psi(\varphi^{-1}(T) \cap \sigma^{-1}(T)) \subseteq T$.

($\Rightarrow$): Assume that $\psi(\varphi^{-1}(T) \cap \sigma^{-1}(T)) \subseteq T$. Let $u, v, w \in \Sigma^*$ be such that $u\ \omega_T\ v$ and $v\ \omega_T\ w$. We wish to show that $u\ \omega_T\ w$. Let $t_1, t_2 \in T$ and $\theta_1, \theta_2 \in \Sigma^*$ be such that $w \in v \sqcup\!\sqcup_{t_1} \theta_1$ and $v \in u \sqcup\!\sqcup_{t_2} \theta_2$. Thus, $w \in (u \sqcup\!\sqcup_{t_2} \theta_2) \sqcup\!\sqcup_{t_1} \theta_1$. Note then that $|t_1|_0 = |t_2|$. Let $n \in \mathbb{N}$ and $\alpha_i, \beta_i \in \mathbb{N}$ be chosen for $1 \leq i \leq n$ so that

$$t_1 = \prod_{i=1}^{n} 0^{\alpha_i} 1^{\beta_i}.$$

Furthermore, let $t_2 = \prod_{i=1}^{n} s_i$ be so that $|s_i| = \alpha_i$ for all $1 \leq i \leq n$. For all $1 \leq i \leq n$, let η_i be the word obtained from s_i by replacing 0 with x and 1 with y, i.e., $\{\eta_i\} = \sigma^{-1}(s_i) \cap \{x, y\}^*$. Then let

$$t = \prod_{i=1}^{n} \eta_i z^{\beta_i}.$$

We can verify that $\varphi(t) = t_1$ and $\sigma(t) = t_2$. Thus, $t \in \varphi^{-1}(t_1) \cap \sigma^{-1}(t_2)$. Let $t' = \psi(t)$. By assumption, $t' \in T$, and we further note that

$$t' = \prod_{i=1}^{n} s_i 1^{\beta_i}.$$

We now define a morphism $h : D^* \to \{0,1\}^*$ given by $h(x) = \epsilon$, $h(y) = 0$ and $h(z) = 1$. Let $\theta \in \theta_2 \sqcup\!\sqcup_{h(t)} \theta_1$. Then we can verify that $w \in (u \sqcup\!\sqcup_{t_2} \theta_2) \sqcup\!\sqcup_{t_1} \theta_1 \subseteq u \sqcup\!\sqcup_{t'} \theta \subseteq u \sqcup\!\sqcup_T \Sigma^*$. Thus, $u\ \omega_T\ w$ as required. ■

Remark 6.4.19 *As an alternate formulation for Theorem 6.4.18, we note that, for all $T \subseteq \{0,1\}^*$, T is transitive if and only if $T \sqcup\!\sqcup_T 1^* \subseteq T$. The reader can verify this by establishing that $T \sqcup\!\sqcup_T 1^* = \psi(\varphi^{-1}(T) \cap \sigma^{-1}(T))$ holds for all $T \subseteq \{0,1\}^*$.*

Corollary 6.4.20 *Given a regular set $T \subseteq \{0,1\}^*$ of trajectories, it is decidable whether T is transitive.*

Proof. Since the regular languages are closed under morphism and inverse morphism, and inclusion of regular languages is decidable, we can determine whether the inclusion (6.9) holds. ■

The following decidability result also holds, since we can determine whether $T \supseteq 0^*$ and (6.9) hold if T is regular:

Corollary 6.4.21 *Given a regular set $T \subseteq \{0,1\}^*$ of trajectories, it is decidable whether ω_T is a partial order.*

We now turn to undecidability. We will use PCP, which we defined in Section 2.5.

Theorem 6.4.22 *Given a CF set $T \subseteq \{0,1\}^*$ of trajectories, it is undecidable whether T is transitive.*

Proof. Let $P = (u_1, u_2, \ldots, u_n; v_1, v_2, \ldots, v_n)$ be a PCP instance. Define

$$\begin{aligned} L_1 &= \{01^{i_1}01^{i_2}\cdots 01^{i_m}0^{n+1}1^{n+1}u_{i_m}u_{i_{m-1}}\cdots u_{i_1} \;:\; m \geq 1, 1 \leq i_p \leq n, 1 \leq p \leq m\}; \\ L_2 &= \{01^{i_1}01^{i_2}\cdots 01^{i_m}0^{n+1}1^{n+1}v_{i_m}v_{i_{m-1}}\cdots v_{i_1} \;:\; m \geq 1, 1 \leq i_p \leq n, 1 \leq p \leq m\}. \end{aligned}$$

Let $K = L_1 \cap L_2$. It is easy to see that P has a solution if and only if $K \neq \emptyset$. Let $T = \{0,1\}^* - K$. Thus, P has no solutions if and only if $T = \{0,1\}^*$. It is easily verified that that T is a CFL.

We now show that P has no solutions if and only if T is transitive. If P has no solutions, then clearly $T = \{0,1\}^*$ is transitive.

Assume that P has a solution. Then there is some word

$$t = 01^{i_1}01^{i_2}\cdots 01^{i_m}0^{n+1}1^{n+1}u_{i_m}u_{i_{m-1}}\cdots u_{i_1} \notin T.$$

Note that $(L_1 \cup L_2) \cap 0^2(0+1)^* = \emptyset$, since $m \geq 1$, and $i_p \geq 1$ for all $1 \leq p \leq m$. Thus, we have that

$$t_1 = 001^{i_1-1}01^{i_2}\cdots 01^{i_m}0^{n+1}1^{n+1}u_{i_m}u_{i_{m-1}}\cdots u_{i_1} \notin K \subseteq L_1 \cup L_2.$$

Thus $t_1 \in T$. Let $\alpha = |t_1|_0$. Note that as $n \geq 1$, certainly $|x|_1 \geq 2$ for all $x \in L_1 \cup L_2$. Thus, we have $t_2 = 010^{\alpha-2} \in T$.

Assume now that T is transitive, contrary to what we want to prove. By (6.9), as $t_1, t_2 \in T$, we must have that $\psi(\varphi^{-1}(t_1) \cap \sigma^{-1}(t_2)) \subseteq T$. But it is easy to verify that $t \in \psi(\varphi^{-1}(t_1) \cap \sigma^{-1}(t_2))$, which is a contradiction. Thus, T is not transitive.

Therefore P has a solution if and only if T is not transitive, and we conclude that it is undecidable whether T is transitive. ■

Consider, by (6.9), or by direct observation, that if $\{T_i\}_{i \in I}$ is a family of transitive sets of trajectories, then the set $\cap_{i \in I} T_i$ is also transitive. Thus, we can define the transitive closure of a set T of trajectories as follows: for all $T \subseteq \{0,1\}^*$, let $tr(T) = \{T' \subseteq \{0,1\}^* \ : \ T \subseteq T', T' \text{ transitive}\}$. Note that $tr(T) \neq \emptyset$, as $\{0,1\}^* \in tr(T)$ for all $T \subseteq \{0,1\}^*$. Define $\widehat{T}$ as

$$\widehat{T} = \bigcap_{T' \in tr(T)} T'. \tag{6.10}$$

Then note that $\widehat{T}$ is transitive and is the smallest transitive set of trajectories containing T. The operation $\widehat{\cdot} : 2^{\{0,1\}^*} \to 2^{\{0,1\}^*}$ is indeed a closure operator (much like the closure operators on sets of trajectories constructed by Mateescu *et al.* [147] for, e.g., associativity and commutativity) in the algebraic sense, since $T \subseteq \widehat{T}$, and $\widehat{\cdot}$ preserves inclusion and is idempotent. Thus, we can, for instance, note the following result:

Lemma 6.4.23 *If* $T \supseteq O$ $(= 0^*1^*0^*)$, *then* $\widehat{T} = H$ $(= \{0,1\}^*)$.

Proof. The result follows, since it is known (and easily observed) that $\widehat{O} = H$ (see, e.g., Ito *et al.* [77, Rem. 3.2]). ■

For particular instances of Lemma 6.4.23, see Thierrin and Yu [193, Prop. 2.3] or Long [136, Thm. 2.1].

A partial order is said to be a *division ordering* [24] if it is positive and compatible.

Lemma 6.4.24 *Let* $T \subseteq \{0,1\}^*$ *be a partial order. If* T *is a division ordering, then* $T = (0+1)^*$.

Proof. As T is positive and compatible, then $T \supseteq 1^*$ and $T \supseteq 0^*T0^*$. Thus, $T \supseteq O$. As T is a partial order, then T is transitive. Thus, $T = \widehat{T} \supseteq \widehat{O} = H$. The result follows. ■

Consider the operator $\Omega_T : 2^{\{0,1\}^*} \to 2^{\{0,1\}^*}$ given by

$$\Omega_T(T') = T \cup T' \cup \psi(\sigma^{-1}(T') \cap \varphi^{-1}(T')). \tag{6.11}$$

By definition of Ω_T, any fixed point $\Omega_T(T_0) = T_0$ contains T and is transitive. Then we have

$$\emptyset \subseteq \Omega_T(\emptyset) = T \subseteq \Omega_T^2(T) \subseteq \Omega_T^3(T) \subseteq \cdots \subseteq \widehat{T}.$$

Since the operations of ϵ-free morphism, inverse morphism, union and intersection are monotone and continuous [158], Ω_T is monotone and continuous and thus $\widehat{T}$ is the least upper bound of $\{\Omega_T^i(\emptyset)\}_{i\geq 0}$. Thus, given T, we can find $\widehat{T}$ by iteratively applying Ω_T to T, and in fact

$$\widehat{T} = \bigcup_{i\geq 0} \Omega_T^i(T). \tag{6.12}$$

This observation allows us to construct $\widehat{T}$, and, for instance, gives us the following result (a similar result for ω-trajectories is given by Kadrie *et al.* [101]):

Lemma 6.4.25 *There exists a regular set of trajectories $T \subseteq \{0,1\}^*$ such that $\widehat{T}$ is not a CFL.*

Proof. Consider $T = (01)^*$, corresponding to perfect or balanced literal shuffle. Then we note that $\widehat{T} \cap 01^* = \{01^{2^n-1} : n \geq 1\}$. ■

Open Problem 6.4.26 *Given $T \in$ REG (or $T \in$ CF), is it decidable whether $\widehat{T} \in$ CF?*

6.4.8 Monotonicity

A binary relation ρ on Σ^* is said to be *monotone* (see, e.g, Ehrenfeucht *et al.* [47, p. 315]) if $x\ \rho\ y$ and $u\ \rho\ v$ implies $xu\ \rho\ yv$ for all $x, y, u, v \in \Sigma^*$. Occasionally, the concept of monotonicity is included as a requirement in compatibility, but we separate the two concepts here for clarity. We note that monotone here is a condition on T, rather than the monotonicity of the operation $⧢_T$

(i.e., that $L_1 \subseteq L_2$, $L_3 \subseteq L_4$, and $T_1 \subseteq T_2$ imply that $L_1 \sqcup\!\sqcup_{T_1} L_3 \subseteq L_2 \sqcup\!\sqcup_{T_2} L_4$), which holds for all T.

Lemma 6.4.27 *Let $T \subseteq \{0,1\}^*$. Then T is monotone if and only if $T^2 \subseteq T$ if and only if $T = T^+$.*

Proof. The fact that $T^2 \subseteq T$ if and only if $T = T^+$ is obvious. Thus, we establish that T is monotone if and only if $T^2 \subseteq T$.

Assume that $T^2 \subseteq T$. Let $x_i \ \omega_T \ y_i$ for $i = 1, 2$. Let $t_i \in T$ and $\alpha_i \in \Sigma^*$ be chosen so that $y_i \in x_i \sqcup\!\sqcup_{t_i} \alpha_i$ for $i = 1, 2$. Then as $t_1 t_2 \in T$, we have the fact that $y_1 y_2 \in x_1 x_2 \sqcup\!\sqcup_{t_1 t_2} \alpha_1 \alpha_2$ implies that $x_1 x_2 \ \omega_T \ y_1 y_2$. Thus T is monotone.

Assume that T is monotone. Let $t_1, t_2 \in T$ be arbitrary. Let $n_i = |t_i|_0$ and $m_i = |t_i|_1$ for $i = 1, 2$. Thus, we have that $0^{n_i} \ \omega_T \ t_i$ for $i = 1, 2$. By the monotonicity of T, $0^{n_1+n_2} \ \omega_T \ t_1 t_2$. Thus, there exist $t \in T$ and $\alpha \in \{0,1\}^*$ such that $t_1 t_2 \in 0^{n_1+n_2} \sqcup\!\sqcup_t \alpha$. But it is now clear that $\alpha = 1^{m_1+m_2}$ and $t = t_1 t_2$. Thus $t_1 t_2 \in T$ and $T^2 \subseteq T$. ■

The following corollary is immediate, since it is decidable whether $T^+ = T$ for regular languages.

Corollary 6.4.28 *Given a regular set T of trajectories, it is decidable whether T is monotone.*

We also have the following undecidability result:

Lemma 6.4.29 *Given an LCF set $T \subseteq \{0,1\}^*$ of trajectories, it is undecidable whether T is monotone.*

Proof. We apply Theorem 2.5.3. First, we note that $T = \{0,1\}^*$ is monotone. Further, we note that the LCF set of trajectories $T = \{0^n 1^n \ : \ n \geq 0\}$ is not expressible as $T = T'/t$ for any monotone $T' \subseteq \{0,1\}^*$ and $t \in \{0,1\}^+$ (Indeed, if this were the case, then as $\epsilon \in T$, $t \in T'$. As $T' = (T')^+$, we have that $t^2, t^3 \in T'$ and $t, t^2 \in T$. But the only way this can happen is if $t = \epsilon$). Thus, we may apply Theorem 2.5.3. and it is undecidable whether a given LCF set of trajectories is monotone. ■

We can now consider the monotone closure of a set T of trajectories, much in the same way we considered the transitive closure in Section 6.4.7. However, we do not need the same level of detail, since it is clear that the monotone closure of T is T^+. Thus, we have the following result:

Lemma 6.4.30 *Let $T \subseteq \{0, 1\}^*$ be a regular (resp., CF, CS, recursive) set of trajectories. Then the monotone closure of T is also a regular (resp., CF, CS, recursive) set of trajectories.*

Recall that $H = \{0, 1\}^*$ and $\mathcal{P}_H$ corresponds to the set of biprefix codes.

Lemma 6.4.31 *Let $T \subseteq \{0, 1\}^*$. If T is ST-strict and monotone, then $\mathcal{P}_T = \mathcal{P}_H$.*

Proof. Let T be ST-strict and monotone. As T is ST-strict, $\epsilon, 0, 1 \in T$. As T is monotone, $\{\epsilon, 0, 1\}^+ = H \subseteq T$. Thus, $T = H$ and the result follows. ■

6.4.9 Well-Foundedness

A partial order ρ is said to be *well-founded* (see, e.g., Choffrut and Karhumäki [24, Sect. 7.1]) if every strictly descending chain under ρ is finite. We note that for relations defined by sets of trajectories, well-foundedness is implied by partial orders (and even by reflexive binary relations):

Theorem 6.4.32 *Let $T \subseteq \{0, 1\}^*$ be a partial order. Then ω_T is a well-founded partial order.*

Proof. Let T be a partial order. Then T is reflexive.

Let $\{w_i\}_{i\geq 1}$ be a descending chain, i.e., $w_{i+1}\ \omega_T\ w_i$ for all $i \geq 1$. Then $|w_{i+1}| \leq |w_i|$ for all $i \geq 1$. Let $K = |w_1|$. Thus, $|w_i| \leq K$ for all $i \geq 1$. Thus, there must exist some $j \geq 1$ such that $|w_j| = |w_{j+1}|$. In particular, this implies that $w_j = w_{j+1}$, and so by the reflexivity of T, $w_j\ \omega_T\ w_{j+1}$. Thus, $\{w_i\}_{i\geq 1}$ is not an infinite strictly descending chain. ■

6.5 Transitivity and Bases

Given a closure operator $\widehat{\cdot}$ and a closed set $S = \widehat{S}$, a base B is a subset $B \subseteq S$ such that $\widehat{B} = S$ and B is minimal with this property with respect to inclusion. Mateescu *et al.* note that in general, problems relating to the existence and effective constructibility of bases are "very challenging mathematically [147, p. 30]." Mateescu *et al.* list several problems relating to bases and associativity for shuffle on trajectories which, to our knowledge, are still open [147, Prob. 3–6, p. 29]. In this section, we investigate the problems of bases with respect to transitivity closure, which we studied in Section 6.4.7.

We will require the following notation. For all $n \geq 1$, let $\vee_n : (\{0,1\}^n)^2 \to \{0,1\}^n$ be pointwise 'OR'. For instance, $0101 \vee_4 1100 = 1101$. Let $\leq_\vee^{(n)}$ be the ordering on the associated poset, i.e., for all $x, y \in \{0,1\}^n$, $x \leq_\vee^{(n)} y$ if and only if $x \vee_n y = y$.

We now consider the notion of a transitivity-base. Given $T \subseteq \{0,1\}^*$ such that T is transitive, a set $B \subseteq \{0,1\}^*$ of trajectories is said to be a *transitivity-base* for T if $B \subseteq T$, $\widehat{B} = T$ and B is minimal with respect to inclusion for the above properties (recall that $\widehat{\cdot}$ is the transitive closure operator defined in Section 6.4.7, cf., (6.10) and (6.12)).

Let $\Pi : 2^{\{0,1\}^*} \to 2^{\{0,1\}^*}$ be defined by

$$\Pi(T) = \psi(\varphi^{-1}(T - 0^*) \cap \sigma^{-1}(T - 0^*)).$$

Note that by Remark 6.4.19, $\Pi(T) = (T - 0^*) \sqcup_{T-0^*} 1^*$. Further, let $\alpha : 2^{\{0,1\}^*} \to 2^{\{0,1\}^*}$ be given by

$$\alpha(T) = T - \Pi(T).$$

We now establish that every language has a transitivity-base.

Theorem 6.5.1 *Let $T \subseteq \{0,1\}^*$ be a transitive set of trajectories. Then $\alpha(T)$ is a transitivity-base for T.*

Proof. Clearly, $\alpha(T) \subseteq T$. Thus, we first demonstrate that $\widehat{\alpha(T)} = T$. Let $t \in T$ be arbitrary. We establish by induction on the length of t that $t \in \widehat{\alpha(T)}$.

For the base case, suppose that t is a trajectory of minimal length in T. Suppose, contrary to what we want to prove, that $t \notin \widehat{\alpha(T)} \supseteq \alpha(T)$. Then as $t \notin \alpha(T)$, $t \in \Pi(T)$. Let $t_1, t_2 \in T - 0^*$ be such that $t \in \psi(\varphi^{-1}(t_1) \cap \sigma^{-1}(t_2))$. By definition of ψ, φ, σ, there exist $n \geq 1$, $i_j, k_j \geq 0$ for $1 \leq j \leq n$ and $s_j \in \{0,1\}^{i_j}$ for $1 \leq j \leq n$ such that

$$\begin{aligned} t_1 &= \prod_{j=1}^{n} 0^{i_j} 1^{k_j} \\ t_2 &= \prod_{j=1}^{n} s_j \\ t &= \prod_{j=1}^{n} s_j 1^{k_j}. \end{aligned}$$

As $\sum_{j=1}^{n} k_j \neq 0$, $t \neq t_2$ and $|t_2| < |t|$. As $t_2 \notin 0^*$, $t \neq t_1$. Now $t_2 \in T$ and $|t_2| < |t|$ contradicts our choice of t.

Assume that for all $t \in T$ with $|t| \leq n$, $t \in \widehat{\alpha(T)}$. Let $m = \min\{n' > n \ : \ T \cap \{0,1\}^{n'} \neq \emptyset\}$. We now establish that for all $t \in T \cap \{0,1\}^m$, $t \in \widehat{\alpha(T)}$. Let $p \geq 1$ and $t_1, t_2, \ldots, t_p$ be the trajectories of length m in T, ordered by $\leq_{\vee}^{(m)}$. We establish the result by induction.

Let t_1 be any trajectory in T of length m which is minimal under $\leq_{\vee}^{(m)}$. Assume that $t_1 \notin \widehat{\alpha(T)}$. Then $t_1 \notin \alpha(T)$ as well. Let $s_1, s_2 \in T - 0^*$ be such that $t_1 \in \psi(\varphi^{-1}(s_1) \cap \sigma^{-1}(s_2))$. Note that $s_1 \neq t_1$, $|s_1| = |t_1|$ and $s_1 \leq_{\vee}^{(m)} t_1$, contradicting our choice of t_1. Thus, $t_1 \in \widehat{\alpha(T)}$.

Now, let $t \in T$ be any word in T of length m, and assume that for all $s \in T$ of length m such that $s \leq_{\vee}^{(m)} t$, $s \in \widehat{\alpha(T)}$ Assume, contrary to what we want to prove, that $t \notin \widehat{\alpha(T)}$. Then again, $t \notin \alpha(T)$ and thus there exist $s_1, s_2 \in T - 0^*$ such that $t \in \psi(\varphi^{-1}(s_1) \cap \sigma^{-1}(s_2))$. Note that $|s_2| < |s_1| = |t|$, and $s_1 \neq t$. Thus, by induction on $|t|$, $s_2 \in \widehat{\alpha(T)}$. By induction on the partial ordering induced by $\leq_{\vee}^{(m)}$, $s_1 \in \widehat{\alpha(T)}$. By Theorem 6.4.18, as $\widehat{\alpha(T)}$ is transitive, $t \in \widehat{\alpha(T)}$. Thus, $T \cap \{0,1\}^m \subseteq \widehat{\alpha(T)}$. Therefore, by induction $T \subseteq \widehat{\alpha(T)}$ and as $\widehat{\cdot}$ preserves inclusion and $T = \widehat{T}$ (T is transitive), the reverse inclusion also holds.

Thus, it remains to establish that $\alpha(T)$ is minimal with respect to inclusions among all T' with $\widehat{T'} = T$. Assume, contrary to what we want to prove, that there exists $T' \subseteq \{0,1\}^*$ such that

$T' \subset \alpha(T)$, where the inclusion noted is proper, and $\widehat{T'} = T$.

Recall the operator Ω_T defined by (6.11). Let $j = \min\{i : \Omega^i_{T'}(T') \cap (\alpha(T) - T') \neq \emptyset\}$. Clearly j exists, as $\emptyset \neq (\alpha(T) - T') \subseteq T = \widehat{T'} = \cup_{i \geq 0} \Omega^i_{T'}(T')$. Let $t \in \Omega^j_{T'}(T')$ be arbitrary. We show that $t \notin \alpha(T) - T'$, contrary to our choice of j. This will give us our contradiction.

Consider that $t \in \Omega^j_{T'}(T') = \Omega^{j-1}_{T'}(T') \cup T' \cup \psi(\varphi^{-1}(\Omega^{j-1}_{T'}(T')) \cap \sigma^{-1}(\Omega^{j-1}_{T'}(T')))$. If $t \in \Omega^{j-1}_{T'}(T')$, then by choice of j, $t \notin \alpha(T) - T'$. Also, if $t \in T'$, then $t \notin \alpha(T) - T'$. Thus, assume that there exist $t_1, t_2 \in \Omega^{j-1}_{T'}(T') \subseteq T$ such that $t \in \psi(\varphi^{-1}(t_1) \cap \sigma^{-1}(t_2))$. Note that if $t_1, t_2 \notin 0^*$, then $t \in \Pi(T)$. In this case, $t \notin \alpha(T)$. Thus, we may assume that $t_1 \in 0^*$ or $t_2 \in 0^*$.

If $t_1 \in 0^*$, then we can see that $t = t_2 \notin \alpha(T) - T'$. Further, if $t_2 \in 0^*$, then we see that $t = t_1 \notin \alpha(T) - T'$. Thus, we have established our contradiction, and $\alpha(T)$ is a transitivity-base for T. ■

We have the following corollary:

Corollary 6.5.2 *Let T be a finite (resp., regular, context-free, recursive) transitive set of trajectories. Then T has a finite (resp., regular, co-*NP*, recursive) transitivity-base.*

Proof. The cases when T is finite, regular or recursive are immediate, based on the closure properties of these classes of languages. We turn to the case when T is a CF set of trajectories. Consider that

$$\Pi(T) = \psi(\varphi^{-1}(T - 0^*) \cap \sigma^{-1}(T - 0^*)).$$

and that $\alpha(T) = T - \Pi(T)$. Note that $T - 0^*$, $\varphi^{-1}(T - 0^*)$ and $\sigma^{-1}(T - 0^*)$ are all CFLs. We claim now that $\Pi(T)$ is in NP. To see this, note that ψ is a letter-to-letter morphism (i.e., $\psi(a) \in \{0, 1\}$ for all $a \in \{x, y, z\}$). Thus, to determine if a trajectory t is a member of $\Pi(T)$, we nondeterministically guess a trajectory t_1 of the same length as t. We then test whether $t \in \psi(t_1)$, and whether $t_1 \in \varphi^{-1}(T - 0^*) \cap \sigma^{-1}(T - 0^*)$. As testing membership in CFLs can be done in polynomial time, and as t_1 is the same length as t, the above is a nondeterministic polynomial-time algorithm for determining membership in $\Pi(T)$. It follows that $\alpha(T) = T - \Pi(T)$ is in co-NP. ■

Example 6.5.3: We give some examples:

(a) let $T = 0^*1^*$. We can compute that $\alpha(T) = 0^*(\epsilon + 1)$.

(b) If $T = (0 + 1)^*$, then $\alpha(T) = 0^*(\epsilon + 1)0^*$.

(c) Let $T = \{0^i 1^{2j} 0^i \ : \ i, j \geq 0\}$. Then $\alpha(T) = (00)^* + \{0^i 110^i \ : \ i \geq 0\}$. Note that T and $\alpha(T)$ are both context-free.

(d) If $T = 1^*0^*1^*$, then $\alpha(T) = 0^* + 10^* + 0^*1$.

□

We now show that the context-free languages are not closed under α. This requires a slightly more complex construction, which we give now:

Theorem 6.5.4 *The context-free languages are not closed under* α.

Proof. Let $T = \{0^i 1^j \ : \ 1 \leq i \leq j\}^*$. We leave it to the reader to verify that $T \in \text{CF}$ and T is transitive. Note that $T \cap 0^* = \emptyset$.

Claim 6.5.5 *If* $t \in \Pi(T)$, *then* $3|t|_0 \leq |t|_1$.

Let $t \in \Pi(T)$, and let $t_1, t_2 \in T$ be such that $t \in t_1 \sqcup\!\sqcup_{t_2} 1^*$. As $t_1, t_2 \in T$, we have $|t_i|_0 \leq |t_i|_1$ for $i = 1, 2$. Further, we note that $|t_1|_0 + |t_1|_1 = |t_1| = |t_2|_0$, $|t|_0 = |t_1|_0$, and $|t|_1 = |t_1|_1 + |t_2|_1$. Thus, we have that

$$3|t|_0 = 3|t_1|_0 \leq |t_1|_1 + 2|t_1|_0 \leq |t_1|_1 + (|t_1|_0 + |t_1|_1) \leq |t_1|_1 + |t_2|_0 \leq |t_1|_1 + |t_2|_1 = |t|_1.$$

Thus, the claim is proven. □

We now return to the proof that $\alpha(T)$ is not a CFL. Assume, contrary to what what we want to prove, that $\alpha(T)$ is a CFL. Then $\alpha(T) \cap 0^+1^+0^+1^+$ is also a CFL.

We employ Ogden's Lemma [68, Lemma 6.2]. Let n be the constant associated with Ogden's Lemma. Assume without loss of generality that $n \geq 1$. Let $t = 0^n 1^n 0^n 1^{5n-1} \in T$. Note that $|t|_0 = 2n$ and $|t|_1 = 6n - 1$, thus $t \notin \Pi(T)$. Therefore, $t \in \alpha(T) \cap 0^+1^+0^+1^+$. Let us consider the first n occurrences of zero as marked. Let $t = uvwxy$. Then we note that both v, x must occur within a block of letters of one type, otherwise, we can consider $t = uv^2wx^2y \notin 0^+1^+0^+1^+$. Now, v or x must contain at least one of the marked letters. Note that if either (a) vwx is entirely contained in the first block of zeroes or (b) v is contained in the first block of zeroes and x is contained in the second block of zeroes or the second block of ones, then uv^2wx^2y has the form $0^{n+k}1^nz$ for some word z starting with zero. This word is clearly not in T, thus not in $\alpha(T)$.

Thus, we must have that v is contained in the first block of zeroes, and x is contained in the first block of ones. Let $v = 0^i$ and $x = 1^j$ for some $i, j \geq 0$ with $i > 0$ and $j \geq 0$. We have two cases:

(a) $i \neq j$. Let $k = 0$ if $i > j$ and $k = 2$ if $i < j$. Then note that $n + (k-1)i > n + (k-1)j$ for this choice of k. Further, $uv^kwx^ky = 0^{n+(k-1)i}1^{n+(k-1)j}0^n1^{5n-1}$. Clearly, $uv^kwx^ky \notin T$.

(b) $i = j$. Note that $n \geq i = j \neq 0$. Thus, consider $uwy = 0^{n-i}1^{n-i}0^n1^{5n-1}$. We claim that $uwy \in \Pi(T)$. Consider $t_1 = 0^{2n-i}1^{2n-i}$ and $t_2 = 0^{n-i}1^{n-i}0^n1^n0^{2n-i}1^{2n+(i-1)}$. Then $uwy \in t_1 \sqcup\!\sqcup_{t_2} 1^*$. It remains to show that $t_1, t_2 \in T - 0^*$. That $t_1, t_2 \notin 0^*$ is easily observed, as $n \neq 0$ and $i \leq n$. Clearly, $t_1 \in T$. Further, as $2n - i < 2n + (i-1)$ for $i > 0$, $t_2 \in T$ as well.

Thus, $\alpha(T)$ is not a CFL, as required. ■

We briefly discuss the problem of bases for monotone sets of trajectories. Recall that the closure operator for monotonicity is T^+. The problem of finding a base for a monotone set of trajectories is therefore a classical problem; we refer the reader to Brzozowski [20]. In particular, we note that the construction $\mu(T) = T - T^2$ gives a base for a monotone set of trajectories T.

6.6 Convexity and Transitivity

Let $\widehat{T}$ again represent the transitive closure of T. We now examine the relationship between T-codes and $\widehat{T}$-codes for arbitrary $T \subseteq \{0,1\}^*$. We call a language $L \subseteq \Sigma^*$ *T-convex* if, for all $y \in \Sigma^*$ and $x, z \in L$, $x \; \omega_T \; y$ and $y \; \omega_T \; z$ implies $y \in L$.

We now characterize when a language is T-convex using shuffle and deletion along trajectories.

Lemma 6.6.1 *Let $T \subseteq \{0,1\}^*$. Then $L \subseteq \Sigma^*$ is T-convex if and only if $(L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*) \subseteq L$.*

Proof. Let L be a T-convex language. Consider an arbitrary word $x \in (L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*)$. Then there exist $y_1, y_2 \in L$ such that $x \in y_1 \sqcup\!\sqcup_T \Sigma^*$ and $x \in y_2 \rightsquigarrow_{\tau(T)} \Sigma^*$. By Lemma 5.8.1, we have that $y_2 \in x \sqcup\!\sqcup_T \Sigma^*$. Thus, $y_1 \, \omega_T \, x \, \omega_T \, y_2$. By the T-convexity of L, $x \in L$. Thus, the inclusion is established.

The reverse implication is similar. Let $(L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*) \subseteq L$. Let $y_1, y_2 \in L$ and $x \in \Sigma^*$ be such that $y_1 \, \omega_T \, x \, \omega_T \, y_2$. Then $x \in y_1 \sqcup\!\sqcup_T \Sigma^*$ and $y_2 \in x \sqcup\!\sqcup_T \Sigma^*$. Again, Lemma 5.8.1 implies that $x \in y_2 \rightsquigarrow_{\tau(T)} \Sigma^*$. Thus, $x \in L$, by our assumed inclusion, and L is T-convex. ∎

Corollary 6.6.2 *Let $T \subseteq \{0,1\}^*$ be reflexive. Then $L \subseteq \Sigma^*$ is T-convex if and only if $(L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*) = L$.*

Proof. We show that if T is reflexive, then for all $L \subseteq \Sigma^*$,

$$L \subseteq (L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*). \tag{6.13}$$

If T is reflexive, then $0^* \subseteq T$ and $(L \sqcup\!\sqcup_T \Sigma^*) \supseteq (L \sqcup\!\sqcup_{0^*} \{\epsilon\}) = L$. Further, if $T \supseteq 0^*$ then $\tau(T) \supseteq i^*$ and $(L \rightsquigarrow_{\tau(T)} \Sigma^*) \supseteq (L \rightsquigarrow_{i^*} \{\epsilon\}) = L$. Thus, we have established (6.13). ∎

We now turn to decidability:

Corollary 6.6.3 *Let $T \subseteq \{0,1\}^*$ be a regular set of trajectories. Given a regular language L, it is decidable whether L is T-convex.*

Proof. As L, T are regular, so are $L \sqcup\!\sqcup_T \Sigma^*$, $L \rightsquigarrow_{\tau(T)} \Sigma^*$ and $(L \sqcup\!\sqcup_T \Sigma^*) \cap (L \rightsquigarrow_{\tau(T)} \Sigma^*)$. Thus, the inclusion in Lemma 6.6.1 is decidable. ∎

We now turn to our main result of this section:

Theorem 6.6.4 *Let Σ be an alphabet and $T \subseteq \{0,1\}^*$. For all languages $L \subseteq \Sigma^+$, the following two conditions are equivalent:*

(i) L is a $\widehat{T}$-code;

(ii) L is a $\widehat{T}$-convex T-code.

Proof. (i) $\Rightarrow$ (ii): Let $L \subseteq \Sigma^+$ be a $\widehat{T}$-code. Then as $T \subseteq \widehat{T}$, L is a T-code as well. Assume that $u \; \omega_{\widehat{T}} \; v \; \omega_{\widehat{T}} \; w$, with $u, w \in L$. As $\widehat{T}$ is transitive, by definition, $u \; \omega_{\widehat{T}} \; w$. Thus, $u = w$, as $u, w \in L$. Now, by the antisymmetry of $\widehat{T}$, $v \; \omega_{\widehat{T}} \; u$ and $u \; \omega_{\widehat{T}} \; v$ imply $v = u \in L$. Thus, L is $\widehat{T}$-convex.
(ii) $\Rightarrow$ (i): Let $L \subseteq \Sigma^+$ be a T-code, as well as being $\widehat{T}$-convex.

Recall the operator Ω_T given by (6.11). Let $T_i = \Omega_T^i(T)$. Then $\widehat{T} = \cup_{i \geq 0} T_i$, by (6.12). We establish (by induction) that L is a T_i-code for all $i \geq 0$. The result will then follow. To see this, assume L is a T_i-code for all $i \geq 0$. Let $x, y \in L$ be such that $x \; \omega_{\widehat{T}} \; y$. Then there exists $t \in \widehat{T}$ such that $y \in x \sqcup\!\sqcup_t z$ for some $z \in \Sigma^*$. As $t \in \widehat{T}$, there exists $i \geq 0$ such that $t \in T_i$, Thus, $x \; \omega_{T_i} \; y$ and then $x = y$, as required.

We now establish by induction on $i \geq 0$ that L is a T_i-code. For $i = 0$, $T_0 = T$. Thus, L is a T-code by assumption.

Let $i > 0$ and assume that L is a T_{i-1}-code. Let $x, y \in L$ be chosen so that $x \; \omega_{T_i} \; y$. Thus, there exist $t \in T_i$ and $z \in \Sigma^*$ such that $y \in x \sqcup\!\sqcup_t z$. We have that $t \in T_i = \Omega_T(T_{i-1}) = T \cup T_{i-1} \cup \psi(\sigma^{-1}(T_{i-1}) \cap \varphi^{-1}(T_{i-1}))$. If $t \in T \cup T_{i-1}$, then, as $y \in x \sqcup\!\sqcup_t z$, by induction $x = y$.

Consider then the case when $t \in \psi(\sigma^{-1}(T_{i-1}) \cap \varphi^{-1}(T_{i-1}))$. Let $t_0, t_1 \in T_{i-1}$ be such that $t \in \psi(\varphi^{-1}(t_0) \cap \sigma^{-1}(t_1))$. By definition of ψ, σ, φ, we know that we can write

$$t_0 = \prod_{k=1}^{n} 0^{i_k} 1^{j_k}$$

for some $n \in \mathbb{N}$ and $i_k, j_k \in \mathbb{N}$ for all $1 \leq k \leq n$, as well as $t_1 = \prod_{k=1}^{n} s_k$ where $|s_k| = i_k$ for all $1 \leq k \leq n$. Further,

$$t = \prod_{k=1}^{n} s_k 1^{j_k}.$$

As $y \in x \sqcup\!\sqcup_t z$, we can write $x = \prod_{k=1}^{n} x_k$, $z = \prod_{k=1}^{n} \alpha_k \beta_k$, where $x_k, \alpha_k, \beta_k \in \Sigma^*$ satisfy $|x_k| = |s_k|_0$, $|\alpha_k| = |s_k|_1$ and $|\beta_k| = j_k$ for all $1 \leq k \leq n$. Further, let $y = \prod_{k=1}^{n} \gamma_k \beta_k$ where $\gamma_k \in x_k \sqcup\!\sqcup_{s_k} \alpha_k$ for all $1 \leq k \leq n$.

Let $\alpha = \prod_{k=1}^{n} \alpha_k$, $\beta = \prod_{k=1}^{n} \beta_k$ and $\gamma = \prod_{k=1}^{n} \gamma_k$. Then we note that

$$\begin{aligned} y &\in \gamma \sqcup\!\sqcup_{t_0} \beta; \\ \gamma &\in x \sqcup\!\sqcup_{t_1} \alpha. \end{aligned}$$

As $t_0, t_1 \in T_{i-1} \subseteq \widehat{T}$, we conclude that $x \; \omega_{T_{i-1}} \; \gamma \; \omega_{T_{i-1}} \; y$, as well as $x \; \omega_{\widehat{T}} \; \gamma \; \omega_{\widehat{T}} \; y$, and thus $\gamma \in L$, by the $\widehat{T}$-convexity of L.

Finally, we note that $\gamma \; \omega_{T_{i-1}} \; y$ implies that $\gamma = y$, as L is a T_{i-1}-code by induction. Similarly, $x \; \omega_{T_{i-1}} \; \gamma$ implies that $\gamma = x$. We conclude that $x = y$ and, since $x, y \in L$ were chosen arbitrarily, L is a T_i-code. ■

Theorem 6.6.4 was known for the case $O = 0^*1^*0^*$, which corresponds to outfix codes, see, e.g., Shyr and Thierrin [185, Prop. 2]. In this case, $\widehat{O} = H = (0+1)^*$, which corresponds to hypercodes. Theorem 6.6.4 was known to Guo *et al.* [57, Prop. 2] in a slightly weaker form for $B = 0^*1^* + 1^*0^*$. In this case, $\widehat{B} = I = 1^*0^*1^*$, and the convexity is with respect to the factor (or subword) ordering. See also Long [136, Sect. 5] for the case of shuffle codes.

6.7 Closure Properties

We now consider the closure properties of $\mathcal{P}_T$.

6.7.1 Closure under Boolean Operations

We note immediately that $\mathcal{P}_T$ is closed under intersection with arbitrary languages, provided the intersection is non-empty:

Lemma 6.7.1 *Let Σ be an alphabet and $T \subseteq \{0,1\}^*$. Let $L_0 \in \mathcal{P}_T(\Sigma)$ and $L_1 \subseteq \Sigma^+$. If $L_0 \cap L_1 \neq \emptyset$, then $L_0 \cap L_1 \in \mathcal{P}_T(\Sigma)$.*

Further, it is clear that $\mathcal{P}_T$ is closed under union if and only if $T \subseteq 0^* + 1^*$.

Lemma 6.7.2 *Let Σ be an alphabet with $|\Sigma| > 1$ and $T \subseteq \{0,1\}^*$. Then $\mathcal{P}_T(\Sigma)$ is closed under union if and only if $T \subseteq 0^* + 1^*$.*

Proof. If $T \subseteq 0^* + 1^*$, then $\mathcal{P}_T(\Sigma) = 2^{\Sigma^+} - \{\emptyset\}$. Thus, it is clear that $\mathcal{P}_T(\Sigma)$ is closed under union.

If $T \cap \overline{0^* + 1^*} \neq \emptyset$, then let $t \in T$ be such that $|t|_0, |t|_1 \neq 0$. Let $t_0 = 0^{|t|_0}$. As $|t|_1 \neq 0$, $t_0 \neq t$. It suffices to note that $\{t\}, \{t_0\} \in \mathcal{P}_T(\Sigma)$, but that $\{t, t_0\} \notin \mathcal{P}_T(\Sigma)$. ■

For completeness, we consider closure of $\mathcal{P}_T(\Sigma)$ under non-empty complement relative to Σ^+:

Lemma 6.7.3 *Let Σ be an alphabet with $|\Sigma| > 1$. Let $T \subseteq \{0,1\}^*$. Then there exists $L \in \mathcal{P}_T(\Sigma)$ such that $\overline{L} \cap \Sigma^+ \neq \emptyset$ and $\overline{L} \cap \Sigma^+ \notin \mathcal{P}_T(\Sigma)$ if and only if $T \not\subseteq 0^* + 1^*$.*

Proof. If $T \subseteq 0^* + 1^*$, then $\mathcal{P}_T(\Sigma) = 2^{\Sigma^+} - \{\emptyset\}$. Assume there exists $L \in \mathcal{P}_T(\Sigma)$ such that $\overline{L} \cap \Sigma^+ \neq \emptyset$ and $\overline{L} \cap \Sigma^+ \notin \mathcal{P}_T(\Sigma)$ Thus, $\overline{L} \cap \Sigma^+ \notin 2^{\Sigma^+} - \{\emptyset\}$. As $\overline{L} \cap \Sigma^+ \neq \emptyset$, we must have that $\overline{L} \cap \Sigma^+ \notin 2^{\Sigma^+}$, i.e., $\overline{L} \cap \Sigma^+ \not\subseteq \Sigma^+$, which is absurd.

If $T \cap \overline{0^* + 1^*} \neq \emptyset$, then let $t \in T$ be such that $|t|_0, |t|_1 \neq 0$. Let $0, 1 \in \Sigma$ without loss of generality.

Let $t_1 = 1^{|t|_1}$ and $t_0 = 0^{|t|_0}$. Note that the three trajectories t, t_0, t_1 are all distinct. We conclude by noting that $L = \{t_1\} \in \mathcal{P}_T(\Sigma)$, but $\Sigma^+ - \{t_1\} \supseteq \{t_0, t\}$. Thus, $L \in \mathcal{P}_T(\Sigma)$ but $\overline{L} \cap \Sigma^+ \notin \mathcal{P}_T(\Sigma)$. ■

6.7.2 Closure under Catenation

Theorem 6.7.4 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that $s(T) \cup p(T) \subseteq T$. Then $\mathcal{P}_T$ is closed under catenation.*

Proof. Let $L_i \in \mathcal{P}_T$ for $i = 1, 2$. Assume that

$$(L_1 L_2 \sqcup\!\sqcup_T x) \cap L_1 L_2 \neq \emptyset$$

for some $x \in \Sigma^*$. We will demonstrate that $x = \epsilon$. Let $\alpha_i, \beta_i \in L_i$ for $i = 1, 2$ be such that

$$\beta_1 \beta_2 \in \alpha_1 \alpha_2 \sqcup\!\sqcup_T x.$$

Let $t \in T$ be such that $\beta_1\beta_2 \in \alpha_1\alpha_2 \sqcup\!\sqcup_t x$. Let $x = x_1x_2$ and $t = t_1t_2$ be chosen so that $\beta_1\beta_2 \in (\alpha_1 \sqcup\!\sqcup_{t_1} x_1)(\alpha_2 \sqcup\!\sqcup_{t_2} x_2)$. We distinguish two cases:

(a) $|\alpha_1| + |x_1| \geq |\beta_1|$. Then there exists $\gamma \in \Sigma^*$ such that

$$\begin{aligned} \beta_1\gamma &\in \alpha_1 \sqcup\!\sqcup_{t_1} x_1; \\ \beta_2 &\in \gamma(\alpha_2 \sqcup\!\sqcup_{t_2} x_2). \end{aligned}$$

Let $t_2' = 1^{|\gamma|}t_2$ and $x_2' = \gamma x_2$. Then, as $|\gamma| \leq |t_1|$, $t_2' \in s(T) \subseteq T$ and thus $\beta_2 \in \alpha_2 \sqcup\!\sqcup_{t_2'} x_2'$ implies that $x_2' = \epsilon$. In particular, $x_2 = \gamma = \epsilon$. As $\gamma = \epsilon$, $\beta_1 \in \alpha_1 \sqcup\!\sqcup_{t_1} x_1$. Note that $t_1 \in p(T) \subseteq T$. Thus, L_1 a T-code implies that $x_1 = \epsilon$ and hence $x = x_1x_2 = \epsilon$.

(b) $|\alpha_1| + |x_1| < |\beta_1|$. Let $\gamma \in \Sigma^+$ be such that

$$\begin{aligned} \beta_1 &\in (\alpha_1 \sqcup\!\sqcup_{t_1} x_1)\gamma; \\ \gamma\beta_2 &\in \alpha_2 \sqcup\!\sqcup_{t_2} x_2. \end{aligned}$$

Let $t_1' = t_1 1^{|\gamma|} \in p(T) \subseteq T$, as $|\gamma| \leq |t_2|$, and let $x_1' = x_1\gamma$. Then $\beta_1 \in (\alpha_1 \sqcup\!\sqcup_{t_1'} x_1')$. As L_1 is a T-code, $x_1' = \epsilon$. This contradicts that $\gamma \in \Sigma^+$.

Thus, $x = \epsilon$ and L_1L_2 is a T-code. ∎

We note that Theorem 6.7.4 can also be proven as follows: as $p(T) \cup s(T) \subseteq T$, T is both cancellative and leviesque. By Jürgensen *et al.* [99, Prop. 10], this implies that $\mathcal{P}_T$ is closed under catenation.

6.7.3 Closure under Inverse Morphism

We now turn to inverse morphism. Let $n \geq 1$. Let $T \subseteq (0^*1^*)^n$ be a bounded regular language such that there exist a_i, b_i, c_i, d_i for $1 \leq i \leq n$ such that

$$T = \prod_{i=1}^{n} 0^{a_i}(0^{b_i})^* 1^{c_i}(1^{d_i})^*. \tag{6.14}$$

(We assume throughout that $T \subseteq (0^*1^*)^n$; similar proofs follow if, e.g., $T \subseteq (0^*1^*)^n 0^*$). Let

$$\begin{aligned} I_j &= \{a_j + b_j m : m \geq 0\} \quad \forall 1 \leq j \leq n; \\ K_j &= \{c_j + d_j m : m \geq 0\} \quad \forall 1 \leq j \leq n. \end{aligned}$$

Let $I'_j = I_j \setminus \{0\}$ for all $1 \leq j \leq n$.

Let $\varphi : \Delta^* \to \Sigma^*$ be a morphism. We define $[\varphi], [\varphi^{-1}] : \mathbb{N} \to 2^{\mathbb{N}}$ as follows:

$$\begin{aligned} [\varphi](m) &= \{|x| : x \in \varphi(\Sigma^m)\}; \\ [\varphi^{-1}](m) &= \{|x| : x \in \varphi^{-1}(\Sigma^m)\}. \end{aligned}$$

We extend these functions naturally to operate on $2^{\mathbb{N}}$ as, e.g., $[\varphi](S) = \bigcup_{s \in S} [\varphi](s)$.

We now prove a generalization of a result on infix and outfix codes established by Ito *et al.* [77, Prop. 6.5].

Theorem 6.7.5 *Let $T \subseteq (0^*1^*)^n$ be a bounded regular set of trajectories as given by (6.14). Let $\varphi : \Delta^* \to \Sigma^*$ be a morphism satisfying*

(a) $\emptyset \neq [\varphi^{-1}](I_j) \subseteq I_j$ for all $1 \leq j \leq n$.

(b) there exists j with $1 \leq j \leq n$ such that $\emptyset \neq [\varphi^{-1}](I'_j) \subseteq I'_j$.

(c) $[\varphi](I_j) \subseteq I_j$ for all $1 \leq j \leq n$.

(d) $[\varphi](K_j) \subseteq K_j$ for all $1 \leq j \leq n$.

Then $\mathcal{P}_T(\Sigma)$ is closed under φ^{-1} if and only if

$$\{|x| : x \in \varphi^{-1}(\epsilon)\}^n \cap \left(\prod_{j=1}^{n} K_j - \{0\}^n \right) = \emptyset. \tag{6.15}$$

Proof. Assume that (6.15) fails. Let x_j for $1 \leq j \leq n$ be such that $x_j \in \varphi^{-1}(\epsilon)$ and $|x_j| \in K_j$. By (6.15), $x = \prod_{i=1}^{n} x_i \neq \epsilon$. Let $k_j = |x_j|$ for $1 \leq j \leq n$.

By (a), let $i_j \in I_j$ be such that $[\varphi^{-1}](i_j) \neq \emptyset$ for all $1 \leq j \leq n$, and such there exist j_0 satisfying $1 \leq j_0 \leq n$, $i_{j_0} \neq 0$ and $[\varphi^{-1}](i_{j_0})$ contains a non-zero element, by (b). Thus, $\varphi^{-1}(\Sigma^{i_j}) \neq \emptyset$. Let

$u_j \in \Sigma^{i_j}$ be such that there exist $v_j \in \varphi^{-1}(u_j)$ for all $1 \leq j \leq n$. As $i_{j_0} \neq 0$, $u = \prod_{j=1}^{n} u_j \neq \epsilon$, and as we can choose $v_{j_0} \in \varphi^{-1}(u_{j_0})$ to be a non-empty word, $v = \prod_{j=1}^{n} v_j \neq \epsilon$. Further, by (a), $|v_j| \in I_j$. Let $\ell_j = |v_j|$ for $1 \leq j \leq n$.

Consider $t = \prod_{j=1}^{n} 0^{\ell_j} 1^{k_j}$. As $\ell_j \in I_j$ and $k_j \in K_j$, $t \in T$. We now define a T-code $L \subseteq \Sigma^+$ such that $\varphi^{-1}(L)$ is not a T-code.

Consider $L = \{u\} \subseteq \Sigma^+$. Trivially, L is a T-code. Let $w = \prod_{j=1}^{n} v_j x_j$. Note that $\varphi(v) = \varphi(v_1) \cdots \varphi(v_n) = u_1 \cdots u_n = u$, and that $\varphi(w) = \prod_{j=1}^{n} \varphi(v_j)\varphi(x_j) = \prod_{j=1}^{n} u_j \cdot \epsilon = u$. Thus, $v, w \in \varphi^{-1}(L)$. Further, $v \neq \epsilon$ implies that $w \neq \epsilon$.

The fact that $\varphi^{-1}(L)$ is not a T-code now follows, since $w \in \varphi^{-1}(L) \cap (v \sqcup\!\sqcup_t x) \subseteq \varphi^{-1}(L) \cap (\varphi^{-1}(L) \sqcup\!\sqcup_T \Delta^+)$.

For the reverse implication, let $L \subseteq \Sigma^+$ be a T-code such that $\varphi^{-1}(L)$ is not a T-code. Then there exists $t \in T$, $u, v \in \varphi^{-1}(L)$ and $x \in \Delta^+$ such that $v \in u \sqcup\!\sqcup_T x$. As $\varphi(u), \varphi(v) \in L \subseteq \Sigma^+$, $u, v \in \Delta^+$.

Consider $t = \prod_{j=1}^{n} 0^{i_j} 1^{k_j}$ for some $i_j \in I_j$ and $k_j \in K_j$ for $1 \leq j \leq n$. Then $v = \prod_{j=1}^{n} u_j x_j$. for $|u_j| = i_j$, $|x_j| = k_j$, $1 \leq j \leq n$. Consider that

$$\begin{aligned} \varphi(v) &= \prod_{i=1}^{n} \varphi(u_j)\varphi(x_j), \\ \varphi(u) &= \prod_{i=1}^{n} \varphi(u_j), \\ \varphi(x) &= \prod_{i=1}^{n} \varphi(x_j). \end{aligned}$$

Let $\ell_j = |\varphi(u_j)|$ and $m_j = |\varphi(x_j)|$ for $1 \leq j \leq n$. By assumptions (c) and (d), $\ell_j \in I_j$ and $m_j \in K_j$. Thus,

$$t' = \prod_{j=1}^{n} 0^{\ell_j} 1^{m_j} \in T.$$

Then we may easily observe that

$$\varphi(v) \in \varphi(u) \sqcup\!\sqcup_{t'} \varphi(x).$$

As $\varphi(v), \varphi(u) \in L$, a T-code, $\varphi(x) = \epsilon$, and, in particular, $\varphi(x_j) = \epsilon$ for all $1 \leq j \leq n$. Thus, recalling that $k_j = |x_j|$ and $x \neq \epsilon$, we note that

$$[k_1, \cdots, k_n] \in \{|x| \; : \; x \in \varphi^{-1}(\epsilon)\}^n \cap \left(\prod_{j=1}^{n} K_j - \{0\}^n \right).$$

This completes the proof. ■

6.7.4 Closure under Reversal

For a word $w = w_1 w_2 \cdots w_n$, where $w_i \in \Sigma$, its *reversal*, denoted w^R, is given by $w^R = w_n w_{n-1} \cdots w_1$. If $L \subseteq \Sigma^*$ is a language, then its reversal is $L^R = \{w^R \; : \; w \in L\}$. For a class of languages $\mathcal{C}$, let $\mathcal{C}^R = \{L^R \; : \; L \in \mathcal{C}\}$.

Lemma 6.7.6 *For all $T \subseteq \{0,1\}^*$, the following equality holds: $\mathcal{P}_{T^R} = \mathcal{P}_T^R$.*

Proof. It suffices to show that $\mathcal{P}_{T^R} \subseteq \mathcal{P}_T^R$.

Let $L \in \mathcal{P}_{T^R}$. Then we have that $L \cap (L \sqcup\!\sqcup_{T^R} \Sigma^+) = \emptyset$. Assume that $L \notin \mathcal{P}_T^R$ and thus $L^R \notin \mathcal{P}_T$. Let $x, y \in L^R$, $t \in T$ and $z \in \Sigma^+$ be such that $x \in y \sqcup\!\sqcup_t z$. Then we note (see, e.g., Mateescu *et al.* [147, Rem. 4.9(ii)]) that $x^R \in y^R \sqcup\!\sqcup_{t^R} z^R$. But as $x^R, y^R \in L$, $t^R \in T^R$, and $z^R \in \Sigma^+$, this contradicts that L is a T^R-code. Thus, $L \in \mathcal{P}_T^R$. ■

Corollary 6.7.7 *Let $T \subseteq \{0,1\}^*$. Then $\mathcal{P}_T^R = \mathcal{P}_T$ if and only if $T = T^R$.*

6.8 Maximal T-codes

Let $T \subseteq \{0,1\}^*$. We say that $L \in \mathcal{P}_T(\Sigma)$ is a *maximal* T-code if, for all $L' \in \mathcal{P}_T(\Sigma)$, $L \subseteq L'$ implies $L = L'$. Denote the set of all maximal T-codes over an alphabet Σ by $\mathcal{M}_T(\Sigma)$. Note that the alphabet Σ is crucial in the definition of maximality. By Zorn's Lemma, we can easily establish that every $L \in \mathcal{P}_T(\Sigma)$ is contained in some element of $\mathcal{M}_T(\Sigma)$. Again, the proof is a specific instance of a result from dependency theory. We may also prove the following result using dependency theory; the result is also clear in our case:

Lemma 6.8.1 *Let $T_1 \subseteq T_2$. Then for all Σ, $\mathcal{M}_{T_2}(\Sigma) \subseteq \mathcal{M}_{T_1}(\Sigma)$.*

6.8.1 Decidability and Maximal T-Codes

Unlike showing that every T-code can be embedded in a maximal T-code, to our knowledge, dependency theory has not addressed the problem of deciding whether a language is a maximal code under some dependence system. We address this problem for T-codes now. We first require the following technical lemma, which is interesting in its own right (specific cases were known for, e.g., prefix codes [18, Prop. 3.1, Thm. 3.3], hypercodes [185, Cor. to Prop. 11], as well as biprefix and outfix codes [134, Lemmas 3.3 and 3.5]). Let $\tau : \{0,1\}^* \to \{i,d\}^*$ be again given by $\tau(0) = i$ and $\tau(1) = d$.

Lemma 6.8.2 *Let $T \subseteq \{0,1\}^*$. Let Σ be an alphabet. For all $L \in \mathcal{P}_T(\Sigma)$, $L \in \mathcal{M}_T(\Sigma)$ if and only if*

$$L \cup (L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+) = \Sigma^+. \tag{6.16}$$

Proof. Let $L \in \mathcal{P}_T(\Sigma) - \mathcal{M}_T(\Sigma)$. Then there exists $x \in \Sigma^+$ such that $L \cup \{x\} \in \mathcal{P}_T(\Sigma)$, but $x \notin L$. Thus, assume, contrary to what we want to prove, that $x \in (L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+)$.

If $x \in L \sqcup\!\sqcup_T \Sigma^+$, then certainly $x \in (L \cup \{x\}) \sqcup\!\sqcup_T \Sigma^+$, by the monotonicity of $\sqcup\!\sqcup_T$. But this contradicts that $L \cup \{x\}$ is a T-code.

If $x \in L \rightsquigarrow_{\tau(T)} \Sigma^+$, then by the monotonicity of $\rightsquigarrow_{\tau(T)}$, $x \in (L \cup \{x\}) \rightsquigarrow_{\tau(T)} \Sigma^+$. But this contradicts that $L \cup \{x\}$ is a T-code, by (6.7). Thus, $x \notin L \cup (L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+)$.

For the reverse implication, assume that $L \in \mathcal{M}_T(\Sigma)$. Then for all $x \in \Sigma^+$ with $x \notin L$, there exist $y \in L, z \in \Sigma^+$ such that either $x \in y \sqcup\!\sqcup_T z$ or $y \in x \sqcup\!\sqcup_T z$. The second membership is equivalent to $x \in y \rightsquigarrow_{\tau(T)} z$. Thus, we have $x \in (L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+)$ for all $x \in \Sigma^+ - L$. The result then follows. ■

Corollary 6.8.3 *Let $T \subseteq \{0,1\}^*$ be a regular set of trajectories. Given a regular language $L \subseteq \Sigma^+$, it is decidable whether $L \in \mathcal{M}_T(\Sigma)$.*

Proof. By Lemma 6.3.9, we can decide whether $L \in \mathcal{P}_T(\Sigma)$. If not, then certainly $L \notin \mathcal{M}_T(\Sigma)$. Otherwise, since T, L are regular, then the languages L, $L \sqcup\!\sqcup_T \Sigma^+$, $L \rightsquigarrow_{\tau(T)} \Sigma^+$, as well as $L \cup$

$(L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+)$ are regular. Thus, the equality (6.16) is decidable. ■

Similar results were also obtained by Kari *et al.* [108, Sect. 5]. We now consider the decidability of being a maximal T-code for finite languages. Our goal is to give a class of sets of trajectories larger than REG such that for any T in our class, it is decidable whether an arbitrary finite language is a maximal T-code.

We first introduce some notation. Let $T \subseteq \{0,1\}^*$. For any $n \geq 0$, let $\eta_n(T) = \{t \in T \ : \ |t|_0 = n\}$. Clearly, $\cup_{n \geq 0} \eta_n(T) = T$.

Before we begin, we require some preliminary lemmas. Recall that a *semilinear set* over $\mathbb{N}^k$ is a finite union of sets of the form $\{\mathbf{u} + \sum_{i=1}^n c_i \mathbf{v_i} \ : \ c_i \in \mathbb{N}\}$ where $\mathbf{u}, \mathbf{v_i} \in \mathbb{N}^k$. The following lemma can be found in Ginsburg [50, Cor. 5.3.2]:

Lemma 6.8.4 *Let $T \subseteq w_1^* w_2^*$ for $w_1, w_2 \in \{0,1\}^*$. Then T is a CFL if and only if $\{(m,n) \ : \ w_1^m w_2^n \in T\}$ is a semilinear set.*

Lemma 6.8.5 *Let $T \subseteq w_1^* w_2^*$ for $w_1, w_2 \in \{0,1\}^*$. If w_1, w_2 are given and T is an effectively given CFL, then for all $n \geq 1$, $\eta_n(T)$ is an effectively regular language.*

For example, let $T = \{0^m 1^m \ : \ m \geq 0\} \subseteq 0^* 1^*$. Then $\eta_n(T) = \{0^n 1^n\}$ for all $n \geq 0$. If $T = (01)^* 1^*$, then $\eta_n(T) = (01)^n 1^*$. We note that we cannot relax the conditions of Lemma 6.8.5 to $T \subseteq w_1^* w_2^* w_3^*$, since, e.g., $T = \{1^n 0^m 1^n \ : \ n, m \geq 0\} \subseteq 1^* 0^* 1^*$, but $\eta_m(T) = \{1^n 0^m 1^n \ : \ n \geq 0\}$, which is not regular if $m > 0$.

Proof. Let $T \subseteq w_1^* w_2^*$ for $w_1, w_2 \in \{0,1\}^*$. Let S be the semilinear set such that $w_1^{\alpha_1} w_2^{\alpha_2} \in T$ if and only if $(\alpha_1, \alpha_2) \in S$. Since the union of regular languages is regular, we can assume without loss of generality that S is linear, i.e., there exist $m, k_1, k_2 \geq 0$ and $p_1, r_i \geq 0$ for all $1 \leq i \leq m$ such that

$$S = \{(k_1, k_2) + \sum_{i=1}^{m} n_i (p_i, r_i) \ : \ (n_1, \ldots, n_m) \in \mathbb{N}^m\}.$$

We assume without loss of generality that $(p_j, r_j) \neq (0,0)$ for all $1 \leq j \leq m$, otherwise, we can simply remove this index from our set without affecting S. We distinguish between four cases:

(a) $w_1 w_2 \in 1^* + 0^*$. In this case, as T is a unary CFL, it is known that T is a regular language. Thus, so is $\eta_n(T) = T \cap (1^*0)^n 1^*$.

(b) $w_1 \in 1^*$. By case (a), we can assume that $w_2 \notin 1^*$, i.e., that $|w_2|_0 \neq 0$. As $w_1 \in 1^*$, there exists $\alpha \geq 0$ such that

$$T = \{1^{\alpha(k_1+\sum_{i=1}^{m} n_i p_i)} w_2^{k_2+\sum_{i=1}^{m} n_i r_i} \ : \ (n_1, \ldots, n_m) \in \mathbb{N}^m\}.$$

Let $I \subseteq \mathbb{N}^m$ be defined so that

$$I = \{(n_1, \ldots, n_m) \ : \ |w_2|_0(k_2 + \sum_{i=1}^{m} n_i r_i) = n\}.$$

From this, we can see that

$$\eta_n(T) = \{1^{\alpha(k_1+\sum_{i=1}^{m} n_i p_i)} w_2^{k_2+\sum_{i=1}^{m} n_i r_i} \ : \ (n_1, \ldots, n_m) \in \mathbb{N}^m\}.$$

By reordering if necessary, let $0 \leq m' \leq m$ be the index such that for all $j \leq m'$, $r_j \neq 0$ and for all $m' < j \leq m$, $r_j = 0$. Let $\varphi : I \to \mathbb{N}^{m'}$ be given by $\varphi(n_1, n_2, \ldots, n_m) = (n_1, n_2, \ldots, n_{m'})$. Note that $\varphi^{-1}(\varphi(I)) = I$ as we have that if $(n_1, \ldots, n_m) \in I$, for all $m' < j \leq m$,

$$(n_1, n_2, \ldots, n_{j-1}, n'_j, n_{j+1}, \ldots, n_m) \in I$$

for all $n'_j \in \mathbb{N}$.

Further, note that $\varphi(I)$ is finite, since for all $(n_1, \ldots, n_{m'}) \in \varphi(I)$ and all $j \leq m'$, n_j satisfies

$$n_j \leq \frac{1}{r_j}\left(\frac{n}{|w_2|_0} - k_2\right).$$

Thus, we can conclude that

$$\eta_n(T) = \{1^{\alpha(k_1+\sum_{i=1}^{m'} n_i p_i)} (\prod_{i=m'+1}^{m} (1^{\alpha p_i})^*) w_2^{k_2+\sum_{i=1}^{m'} n_i r_i} \ : \ (n_1, \ldots, n_{m'}) \in \varphi(I)\}.$$

and that $\eta_n(T)$ is regular.

(c) $w_2 \in 1^*$. Thus, consider that $\eta_n(T^R) = \eta_n(T)^R$. As $T^R \subseteq (w_2^R)^*(w_1^R)^*$, by (a) or (b), $\eta_n(T^R)$ is regular. As the regular languages are closed under reversal, $\eta_n(T)$ is regular.

(d) $w_1, w_2 \notin 1^*$. Let $I \subseteq \mathbb{N}^m$ be defined by

$$I = \{(n_1, \ldots, n_m) \in \mathbb{N}^m \\ : |w_1|_0(k_1 + \sum_{i=1}^{m} n_i p_i) + |w_2|_0(k_2 + \sum_{i=1}^{m} n_i r_i) = n\}.$$

Note that I is finite, as $|w_1|_0, |w_2|_0 \neq 0$ and $(p_i, r_i) \neq (0, 0)$ for all $1 \leq i \leq m$. Further, we have that

$$\eta_n(T) = \{w_1^{k_1 + \sum_{i=1}^{m} n_i p_i} w_2^{k_2 + \sum_{i=1}^{m} n_i r_i} : (n_1, \ldots, n_m) \in I\}.$$

From this, we note that $\eta_n(T)$ is finite.

Thus, $\eta_n(T)$ is regular. ■

We are now ready to give our positive decidability result:

Theorem 6.8.6 *Let $T \subseteq \{0, 1\}^*$ be a CFL such that $T \subseteq w_1^* w_2^*$ for $w_1, w_2 \in \{0, 1\}^*$, where w_1, w_2 are given. If F is a finite set, then we can decide whether F is a maximal T-code. Furthermore, all constructions are effective.*

Proof. Let $T \subseteq w_1^* w_2^*$ be a CFL. Let F be our finite set and let $\ell(F) = \{|x| : x \in F\}$ and $\ell_F = \max\{\ell : \ell \in \ell(F)\}$. First, we note that we can find $T^{\leq \ell_F} = T \cap \{0, 1\}^{\leq \ell_F}$, and that

$$F \rightsquigarrow_{\tau(T)} \Sigma^+ = F \rightsquigarrow_{\tau(T^{\leq \ell_F})} \Sigma^+,$$

which is thus a regular language, since F, Σ^+, $\tau(T^{\leq \ell_F})$ are, as well.

Second, we note that $\eta(T) = \cup_{\ell \in \ell(F)} \eta_\ell(T)$ is a regular language, since $\ell(F)$ is finite, and $\eta_\ell(T)$ is regular by Lemma 6.8.5. Further, we note that

$$F \sqcup_T \Sigma^+ = F \sqcup_{\eta(T)} \Sigma^+,$$

which is regular, by the regularity of F, Σ^+ and $\eta(T)$.

Thus, we conclude that $F \cup (F \rightsquigarrow_{\tau(T)} \Sigma^+) \cup (F \sqcup\!\sqcup_T \Sigma^+)$ is a regular language, and thus, we can determine whether this language is equal to Σ^+. Thus, by Lemma 6.8.2, we can determine whether F is a maximal T-code. ■

6.8.2 Transitivity and Embedding T-codes

Given a class of codes $\mathcal{C}$, and a language $L \in \mathcal{C}$ of given complexity, there has been much research into whether or not L can be *embedded in* (or *completed to*) a maximal element $L' \in \mathcal{C}$ of the same complexity, i.e., a maximal code $L' \in \mathcal{C}$ with $L \subseteq L'$. Finite and regular languages in these classes of codes are of particular interest. For instance, we note that every regular code can be completed to a maximal regular code, while the same is not true for finite codes or finite biprefix codes.

We now show an interesting result on embedding T-codes in maximal T-codes while preserving complexity. For example, we will show that if T is transitive and regular and L is a regular T-code, then we can embed L in a maximal T-code which is also regular.

Our construction is a generalization of a result due to Lam [128]. In particular, we define two transformations on languages. Let T be a set of trajectories and $L \subseteq \Sigma^+$ be a language. Then define $U_T(L)$, $V_T(L) \subseteq \Sigma^+$ as

$$\begin{aligned} U_T(L) &= \Sigma^+ - (L \sqcup\!\sqcup_T \Sigma^+ \cup L \rightsquigarrow_{\tau(T)} \Sigma^+); \\ V_T(L) &= U_T(L) - (U_T(L) \sqcup\!\sqcup_T \Sigma^+). \end{aligned}$$

First, we note the following two properties of $U_T(L)$, $V_T(L)$:

Lemma 6.8.7 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories and $L \in \mathcal{P}_T(\Sigma)$. Then $L \subseteq U_T(L)$ and $L \subseteq V_T(L)$.*

Proof. We establish first that $L \subseteq U_T(L)$. Let $x \in L$, but assume that $x \notin U_T(L)$. Then $x \in L \sqcup\!\sqcup_T \Sigma^+$ or $x \in L \rightsquigarrow_{\tau(T)} \Sigma^+$. In the first case, we have $L \cap (L \sqcup\!\sqcup_T \Sigma^+) \neq \emptyset$, contradicting that L is a T-code. The second case also contradicts that L is a T-code, since then $L \cap (L \rightsquigarrow_{\tau(T)} \Sigma^+) \neq \emptyset$, contradicting (6.7).

We now establish $L \subseteq V_T(L)$. Assume not, then as $L \subseteq U_T(L)$, we must have that $L \cap (U_T(L) \sqcup\!\sqcup_T \Sigma^+) \neq \emptyset$. Assume that $y \in U_T(L)$, $z \in \Sigma^+$ and $x \in L$ are chosen so that $x \in y \sqcup\!\sqcup_T z$. Thus $y \in x \rightsquigarrow_{\tau(T)} z \subseteq L \rightsquigarrow_{\tau(T)} \Sigma^+$, contradicting that $y \in U_T(L)$. Thus, $L \subseteq V_T(L)$. ■

Theorem 6.8.8 *Let $T \subseteq \{0,1\}^*$ be transitive. Let Σ be an alphabet. Then for all $L \in \mathcal{P}_T(\Sigma)$, the language $V_T(L)$ contains L and $V_T(L) \in \mathcal{M}_T(\Sigma)$.*

Proof. By Lemma 6.8.7, $L \subseteq V_T(L)$. That $V_T(L)$ is a T-code follows from Lemma 6.3.11 applied to $U_T(L)$. Thus, it remains to show that for all $z \in \Sigma^+ - V_T(L)$, $V_T(L) \cup \{z\}$ is not a T-code.

Let $z \notin V_T(L)$ be arbitrary. We distinguish two cases:

(a) if $z \notin U_T(L)$, then $z \in (L \sqcup\!\sqcup_T \Sigma^+) \cup (L \rightsquigarrow_{\tau(T)} \Sigma^+)$. If $z \in L \sqcup\!\sqcup_T \Sigma^+ \subseteq V_T(L) \sqcup\!\sqcup_T \Sigma^+$, then $V_T(L) \cup \{z\} \notin \mathcal{P}_T(\Sigma)$. If $z \in L \rightsquigarrow_{\tau(T)} \Sigma^+ \subseteq V_T(L) \rightsquigarrow_{\tau(T)} \Sigma^+$, then again (this time by (6.7)), $V_T(L) \cup \{z\} \notin \mathcal{P}_T(\Sigma)$.

(b) if $z \in U_T(L) - V_T(L)$, then $z \in U_T(L) \sqcup\!\sqcup_T \Sigma^+$. Let $y \in U_T(L)$ be a shortest word such that $z \in y \sqcup\!\sqcup_T \Sigma^+$. We claim that $y \in V_T(L)$. If this were not the case, then as $y \in U_T(L) - V_T(L)$, we have that $y \in U_T(L) \sqcup\!\sqcup_T \Sigma^+$, by definition of $V_T(L)$. Let $y' \in U_T(L)$ be such that $y \in y' \sqcup\!\sqcup_T \Sigma^+$. Thus, we have that $y' \omega_T y \omega_T z$. By transitivity of T, $y' \omega_T z$, i.e., $z \in y' \sqcup\!\sqcup_T \Sigma^*$. As $|y'| < |y| < |z|$, we certainly have that $z \in y' \sqcup\!\sqcup_T \Sigma^+$ in particular. But as $|y'| < |y|$, this contradicts our choice of y. Thus, $y \in V_T(L)$. But $y, z \in V_T(L) \cup \{z\}$ and $z \in y \sqcup\!\sqcup_T \Sigma^+$ imply that $V_T(L) \cup \{z\} \notin \mathcal{P}_T(\Sigma)$.

Thus, $V_T(L)$ is a maximal T-code. ■

There are several consequences of Theorem 6.8.8. We note only one important corollary:

Corollary 6.8.9 *Let $T \subseteq \{0,1\}^*$ be transitive and regular. Then every regular (resp., recursive) T-code is contained in a maximal regular (resp., recursive) T-code.*

Corollary 6.8.9 was given for $T = 1^*0^*1^*$ and regular T-codes by Lam [128, Prop. 3.2]. Further research into the case when T is not transitive is necessary (for example, the proofs of Zhang and Shen [206] and Bruyère and Perrin [19] on embedding regular biprefix codes are much more involved than our construction, and do not seem to be easily generalized).

We can extend our embedding results to finite languages with one additional constraint on T, namely completeness. The following technical lemma is easily proven:

Lemma 6.8.10 *Let $T \subseteq \{0,1\}^*$ be complete. Then for all $y \in \Sigma^*$ and for all $m \leq |y|$, there exists $z \in \Sigma^m$ such that $y \in z \sqcup\!\sqcup_T \Sigma^*$. Further, if $m < |y|$, $y \in z \sqcup\!\sqcup_T \Sigma^+$.*

We now show that for transitive and complete sets of trajectories T, finite T-codes can be completed to finite maximal T-codes.

Corollary 6.8.11 *Let $T \subseteq \{0,1\}^*$ be transitive and complete. Let Σ be an alphabet. Then for all finite $F \in \mathcal{P}_T(\Sigma)$, there exists a finite language $F' \in \mathcal{M}_T(\Sigma)$ such that $F \subseteq F'$. Further, if T is effectively regular, and F is effectively given, we can effectively construct F'.*

Proof. Let F be a finite language and $n = \max\{|x| : x \in F\}$. As $F \in \mathcal{P}_T(\Sigma)$, $n \neq 0$. We first establish the following claim: for all $y \in \Sigma^+$ with $|y| > n$, there exists $u \in U_T(F)$ such that $y \in u \sqcup\!\sqcup_T \Sigma^+$.

Let $y \in \Sigma^+$ be such that $|y| > n$. Then by Lemma 6.8.10, there exists z such that $|z| = n$ and $y \in z \sqcup\!\sqcup_T \Sigma^+$. Note that as $n \neq 0$, $z \in \Sigma^+$. If $z \in U_T(F)$, we have established the claim with $u = z$. Thus, assume that $z \notin U_T(F)$. By definition of $U_T(F)$, we have that $z \in (F \sqcup\!\sqcup_T \Sigma^+) \cup (F \rightsquigarrow_{\tau(T)} \Sigma^+)$. However, $|x| < n$ for all $x \in F \rightsquigarrow_{\tau(T)} \Sigma^+$. Thus, we have that $z \in F \sqcup\!\sqcup_T \Sigma^+ \subseteq U_T(F) \sqcup\!\sqcup_T \Sigma^+$, the inclusion being valid by Lemma 6.8.7. Let $u \in U_T(F)$ be such that $z \in u \sqcup\!\sqcup_T \Sigma^+$. Then $u \, \omega_T \, z$ and $z \, \omega_T \, y$. Thus, by transitivity, $u \, \omega_T \, y$. As $|u| < |y|$, this implies that $y \in u \sqcup\!\sqcup_T \Sigma^+$. Thus, our claim is proven.

We now establish that $V_T(F)$ is finite. Let y be an arbitrary word such that $|y| > n$. By our claim, $y \in U_T(F) \sqcup\!\sqcup_T \Sigma^+$. But by definition of $V_T(F)$, this implies that $y \notin V_T(F)$. Thus, $V_T(F) \subseteq \Sigma^{\leq n}$. Therefore, the conditions of the corollary are met by $V_T(F)$, by Theorem 6.8.8. This completes the proof. ■

In practice, the condition that T be complete is not very restrictive, since natural operations seem to typically be defined by a complete set of trajectories.

In Section 6.9.3 below, we will give alternate conditions on T that ensure that every finite T-code can be embedded in a finite maximal T-code. However, this result will be a trivial consequence of the fact that for such T, all T-codes are finite.

We now show that there exist T which are not transitive, and for which the above results do not hold. It is known, for example, that there exist finite biprefix codes which cannot be embedded in a maximal finite biprefix code (see, e.g., Bruyère and Perrin [19, Sect. 3]). We present the following two examples, as well; in the first case, T is regular but not transitive, and for all regular T-codes L, L cannot be embedded in any maximal CF T-code. In the second example, T is not complete, and no finite T-code can be embedded in a maximal finite T-code.

Example 6.8.12: Let $T = (01)^*$; then $\sqcup\!\sqcup_T$ is known as perfect or balanced literal shuffle. Clearly, T is not transitive. Let $\Sigma = \{a\}$. We claim that for all regular languages $L \subseteq a^*$, L is not a maximal T-code.

Let $L \subseteq a^*$ be regular. As L is a unary regular language, it is well-known that L corresponds to an ultimately periodic set of natural numbers. That is, there exist $n_0, p \in \mathbb{N}$ with $p > 0$ such that for all $n > n_0$, $a^n \in L$ if and only if $a^{n+p} \in L$.

Let $r = \min\{kp : k \geq 1, kp > n_0\}$. Then we have two cases:

(a) if $a^r \in L$, then $a^{2r} \in L$ as well. Thus, as $a^{2r} \in a^r \sqcup\!\sqcup_T a^r$, L is not a T-code.

(b) if $a^r \notin L$, then $a^{2r} \notin L$ as well. Thus, consider $L \cup \{a^{2r}\}$. If L is a T-code, then as $a^{2r} \notin L \sqcup\!\sqcup_T a^+$ and $L \cap (a^{2r} \sqcup\!\sqcup_T a^+) = \emptyset$, we have that $L \cup \{a^{2r}\}$ is a T-code as well. Thus, L is not a maximal T-code.

Thus, there are no regular languages in $\mathcal{M}_T(\{a\})$. Further, since the unary context-free and unary regular languages coincide, there are no context-free languages in $\mathcal{M}_T(\{a\})$, either. Thus, e.g., the T-code $\{a\}$ cannot be embedded in any regular (or context-free) maximal T-code.

We note in passing that one maximal T-code containing $\{a\}$ is given by $L = \{a^{c_n} : n \geq 1\}$

where $\{c_n\}_{n\geq 1} = \{1, 3, 4, 5, 7, 9, 11, \dots\}$ is the lexicographically least sequence of positive integers satisfying $m \in \{c_n\} \iff 2m \notin \{c_n\}$. This sequence has received some attention in the literature, and has connections to the Thue-Morse word. We point the reader to A003159 in Sloane [188] for details and references. Clearly, L is not regular. □

Example 6.8.13: Let $T = \{0^j 1^{2i} 0^j \ : \ i, j \geq 0\}$. Then $\sqcup\!\sqcup_T$ is the balanced insertion operation. Note that T is transitive, but not complete. Let Σ be an alphabet and let $L_o = \{x \in \Sigma^+ \ : \ |x| \equiv 1 \pmod 2\}$. Then for all $L \in \mathcal{P}_T(\Sigma)$, $L \cup L_o \in \mathcal{P}_T(\Sigma)$. Thus, there are no finite maximal T-codes. □

6.9 Finiteness of all T-codes

In this section, we investigate $T \subseteq \{0, 1\}^*$ such that all $\mathcal{P}_T$ codes are finite. It is a well-known result that all hypercodes ($T = \{0, 1\}^*$) are finite, which can be concluded from a result due to Higman [64].

We define the following classes of sets of trajectories:

$$\begin{aligned} \mathfrak{F}_R &= \{T \in \{0, 1\}^* \ : \ \mathcal{P}_T \cap \text{REG} \subseteq \text{FIN}\}; \\ \mathfrak{F}_C &= \{T \in \{0, 1\}^* \ : \ \mathcal{P}_T \cap \text{CF} \subseteq \text{FIN}\}; \\ \mathfrak{F}_H &= \{T \in \{0, 1\}^* \ : \ \mathcal{P}_T \subseteq \text{FIN}\}. \end{aligned}$$

The class $\mathfrak{F}_H$ is of particular importance. If T is a partial order and $T \in \mathfrak{F}_H$, then T is a *well partial order*[1]. This is a subject of tremendous research, not only in the larger theory of partial orders (see the survey of Kruskal [125]), but also within formal language theory as well. Without trying to be exhaustive, we note the work of Jullien [96], Haines [58], van Leeuwen [197], Ehrenfeucht *et al.* [47], Ilie [72, 73], Ilie and Salomaa [80] and Harju and Ilie [59] on well partial orders relating to words. We also refer the reader to the survey of results presented by de Luca and Varricchio [33, Sect. 5].

[1] Recall that we say that T has property P if and only if ω_T has property P.

To begin, we give conditions on T which ensure all regular (or context-free) T-codes are finite.

6.9.1 Finiteness of Regular T-codes

Let $T \subseteq \{0,1\}^*$. Define the *insertion behaviour* of T, denoted $ib(T)$, as

$$ib(T) = \{(n_1, n_2, n_3) \in \mathbb{N}^3 \ : \ 0^{n_1} 1^{n_2} 0^{n_3} \in T\}.$$

Say that T is REG-*pumping compliant* if, for all $i, j, k \in \mathbb{N}$ ($j > 0$), there exists j' with $0 \leq j' < j$ such that

(i) if $j' = 0$, then $ib(T) \cap \{(i + jm_1, jm_2, k + jm_3) \ : \ m_1, m_3 \geq 0, m_2 > 0\} \neq \emptyset$.

(ii) if $1 \leq j' < j$, then $ib(T) \cap \{(i + j' + jm_1, jm_2, k - j' + jm_3) \ : \ m_1 \geq 0, m_2, m_3 > 0\} \neq \emptyset$.

The use of the terminology 'REG-pumping compliant' will become clear in the following lemma:

Lemma 6.9.1 *Let $T \subseteq \{0,1\}^*$. If T is* REG*-pumping compliant, then $T \in \mathfrak{F}_R$.*

Proof. Let $R \in$ REG be an infinite regular language over Σ. By the pumping lemma for regular languages, there exist $u, v, w \in \Sigma^*$ such that $v \neq \epsilon$ and $uv^*w \subseteq R$. Let $i = |u|$, $j = |v|$ and $k = |w|$. Note that $j \neq 0$. Let j' be the natural number implied by the REG-pumping compliance condition.

If $j' = 0$, then let m_1, m_2, m_3 be chosen so that $m_1, m_3 \geq 0$, $m_2 > 0$ and $(i + jm_1, jm_2, k + jm_3) \in ib(T)$. Let $t = 0^{i+jm_1} 1^{jm_2} 0^{k+jm_3}$. By definition, $t \in T$. Consider $x = uv^{m_1+m_3}w \in R$ and $y = v^{m_2}$. As $m_2 \neq 0$ and $v \neq \epsilon$, $y \neq \epsilon$. We note that

$$x \sqcup\!\sqcup_t y \ni uv^{m_1} \cdot v^{m_2} \cdot v^{m_3} w = uv^{m_1+m_2+m_3} w.$$

Thus, $(R \sqcup\!\sqcup_T \Sigma^+) \cap R \neq \emptyset$ and $R \notin \mathcal{P}_T$.

If $1 \leq j' < j$, let $m_1 \geq 0$, $m_2, m_3 > 0$ be chosen so that

$$(i + j' + jm_1, jm_2, k - j' + jm_3) \in ib(T),$$

and hence $t = 0^{i+j'+jm_1}1^{jm_2}0^{k+(j-j')+j(m_3-1)} \in T$. Let $v_1 \in \Sigma^*$ be the prefix of v of length j' and let $v = v_1 v_2$ for some $v_2 \in \Sigma^*$.

Consider $x = uv^{m_1+m_3}w \in R$ and $y = (v_2 v_1)^{m_2} \neq \epsilon$. Then

$$x \sqcup\!\sqcup_t y \ni uv^{m_1}v_1 \cdot v_2(v_1 v_2)^{m_2-1}v_1 \cdot v_2 v^{m_3-1}w = uv^{m_1+m_2+m_3}w.$$

Again, $(R \sqcup\!\sqcup_T \Sigma^+) \cap R \neq \emptyset$ and thus $R \notin \mathcal{P}_T$. Thus, $\mathcal{P}_T$ contains no infinite regular languages. ■

The condition of being REG-pumping compliant is not very restrictive. Clearly, if $T \supseteq 0^*1^*0^*$, then T is REG-pumping compliant (in this case, Lemma 6.9.1 is a corollary of a result on outfix codes due to Ito *et al.* [77]). For a broader class of examples, we can consider immune languages.

Lemma 6.9.2 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that $\overline{T} \cap 0^*1^*0^*$ is* REG-*immune. Then T is* REG-*pumping compliant.*

Proof. Let $i \geq 0$, $j > 0$, $k \geq 0$ be arbitrary. Consider

$$T_0 = T_0(i,j,k) = 0^i(0^j)^*(1^j)^+(0^j)^*0^k.$$

As T_0 is an infinite regular language, T_0 is not a subset of $\overline{T} \cap 0^*1^*0^*$. Thus, $T_0 \cap \overline{(\overline{T} \cap 0^*1^*0^*)} = T_0 \cap (T \cup \overline{0^*1^*0^*}) \neq \emptyset$. As $T_0 \subseteq 0^*1^*0^*$, this implies that $T_0 \cap T \neq \emptyset$. Thus, there exist $m_1 \geq 0$, $m_2 > 0$ and $m_3 \geq 0$ such that $0^{i+jm_1}1^{jm_2}0^{k+jm_3} \in T$, i.e., $(i+jm_1, jm_2, k+jm_3) \in ib(T)$. Thus, the REG-pumping compliant conditions are met with $j' = 0$. ■

Next, we show that if $T \subseteq 0^*1^*0^*$, then REG-pumping compliance is necessary to ensure that there are no infinite regular languages in $\mathcal{P}_T$.

Lemma 6.9.3 *Let $T \subseteq 0^*1^*0^*$ be not* REG-*pumping compliant. Then $\mathcal{P}_T(\Sigma)$ contains an infinite regular language for all Σ with $|\Sigma| \geq 2$.*

Proof. Let $i, j, k \in \mathbb{N}$ be arbitrary such that $i \geq 0$, $j > 0$, $k \geq 0$,

$$ib(T) \cap \{(i+jm_1, jm_2, k+jm_3) \;:\; m_1, m_3 \geq 0, m_2 > 0\} = \emptyset.$$

and for all $1 \le j' < j$,

$$ib(T) \cap \{(i + j' + jm_1, jm_2, k - j' + jm_3) \ : \ m_1 \ge 0, m_2, m_3 > 0\} = \emptyset.$$

Let $a, b \in \Sigma$ $(a \neq b)$ and $R = a^i (b^j)^* a^k$. We claim that $R \in \mathcal{P}_T(\Sigma)$. Assume not. Then there exist $\ell_1 > \ell_2 \ge 0$ such that

$$a^i b^{j\ell_1} a^k \in a^i b^{j\ell_2} a^k \sqcup\!\sqcup_T z$$

for some $z \in \{a, b\}^+$. By observation, $z = b^{j(\ell_1 - \ell_2)}$. Thus, let $t \in T$ be chosen so that

$$a^i b^{j\ell_1} a^k \in a^i b^{j\ell_2} a^k \sqcup\!\sqcup_t b^{j(\ell_1 - \ell_2)}.$$

Then as $T \subseteq 0^*1^*0^*$, $t = 0^{i+\alpha j+j'} 1^{j(\ell_1-\ell_2)} 0^{(j-j')+(\ell_2-\alpha-1)j+k}$ for some α and j' with either $0 \le \alpha \le \ell_2$ and $j' = 0$ or $0 \le \alpha < \ell_2 - 1$ and $1 \le j' < j$. If $j' = 0$, then $(i + \alpha j, j(\ell_1 - \ell_2), k + (\ell_2 - \alpha)j) \in ib(T)$ while if $j' \neq 0$, then $(i + j' + \alpha j, j(\ell_1 - \ell_2), k - j' + (\ell_2 - \alpha)j) \in ib(T)$, which are both contradictions. ■

6.9.2 Finiteness of Context-free T-codes

Let $T \subseteq \{0, 1\}^*$. Define the *2–insertion behaviour* of T, denoted $2ib(T)$, as follows:

$$2ib(T) = \{(n_1, n_2, \ldots, n_5) \in \mathbb{N}^5 \ : \ 0^{n_1} 1^{n_2} 0^{n_3} 1^{n_4} 0^{n_5} \in T\}.$$

We use $2ib(T)$ to define the notion of CF-*pumping compliance*. The idea is the same as REG-pumping compliance, but with more cases. In particular, say that T is CF-pumping compliant if, for all $i, j_1, j_2, k, \ell \in \mathbb{N}$, with $j_1 + j_2 > 0$, there exist $j_1', j_2' \in \mathbb{N}$ such that $0 \le j_i' < j_i$ for $i = 1, 2$ and $2ib(T) \cap P \neq \emptyset$, where P is defined as follows:

(a) if $j_1' = j_2' = 0$, then

$$\begin{aligned} P \ = \ & \{(i + j_1\alpha_1, j_1\beta, k + j_1\alpha_2 + j_2\alpha_3, j_2\beta, \ell + j_2\alpha_4) \\ & : \ \alpha_m, \beta \in \mathbb{N}, (1 \le m \le 4), \beta > 0, \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4\}. \end{aligned}$$

(b) if $1 \le j_1' < j_1$ and $j_2' = 0$, then

$$P = \{(i + j_1' + j_1\alpha_1, j_1\beta, k - j_1' + j_1\alpha_2 + j_2(\alpha_3 + \gamma_1), j_2\beta, \ell + j_2(\alpha_4 + \gamma_2)) \\ : \alpha_m, \beta, \gamma_p \in \mathbb{N}, (1 \le m \le 4, 1 \le p \le 2), \\ \beta, \alpha_2 > 0, \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4 + 1, \gamma_1 + \gamma_2 = 1\}.$$

(c) if $j_1' = 0$ and $1 \le j_2' < j_2$, then

$$P = \{(i + j_1(\alpha_1 + \gamma_1), j_1\beta, k + j_2' + j_1(\alpha_2 + \gamma_2) + j_2\alpha_3, j_2\beta, \ell - j_2' + j_2\alpha_4) \\ : \alpha_m, \beta, \gamma_p \in \mathbb{N}, (1 \le m \le 4, 1 \le p \le 2), \\ \beta, \alpha_4 > 0, \alpha_1 + \alpha_2 + 1 = \alpha_3 + \alpha_4, \gamma_1 + \gamma_2 = 1\}.$$

(d) if $1 \le j_1' < j_1$ and $1 \le j_2' < j_2$, then

$$P = \{(i + j_1' + j_1\alpha_1, j_1\beta, k - j_1' + j_2' + j_1\alpha_2 + j_2\alpha_3, j_2\beta, \ell - j_2' + j_2\alpha_4) \\ : \alpha_m, \beta \in \mathbb{N}, (1 \le m \le 4), \ \beta, \alpha_2, \alpha_4 > 0, \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4\}.$$

Lemma 6.9.4 *Let $T \subseteq \{0, 1\}^*$. If T is* CF*-pumping compliant, then $T \in \mathfrak{F}_C$.*

Proof. Let $L \in$ CF be an infinite language which is a subset of Σ^+. Then by the pumping lemma for CFLs, there exist $u, v, w, x, y \in \Sigma^*$ such that $vx \neq \epsilon$ and $\{uv^m wx^m y : m \ge 0\} \subseteq L$. Let $i = |u|$, $j_1 = |v|$, $k = |w|$, $j_2 = |x|$ and $\ell = |y|$. Let j_1', j_2' be the natural numbers implied by the CF-pumping compliance of T. We consider the case $j_1' = 0$ and $1 \le j_2' < j_2$. The other cases are similar (the differences are similar to the differences between the cases in the proof of Lemma 6.9.1).

Let $\alpha_m, \beta, \gamma_p \in \mathbb{N}$ for $1 \le m \le 4$ and $1 \le p \le 2$ be such that

$$(i + j_1(\alpha_1 + \gamma_1), j_1\beta, k + j_2' + j_1(\alpha_2 + \gamma_2) + j_2\alpha_3, j_2\beta, \ell - j_2' + j_2\alpha_4) \in 2ib(T). \tag{6.17}$$

Further, we have that $\beta, \alpha_4 > 0$, $\alpha_1 + \alpha_2 + 1 = \alpha_3 + \alpha_4$ and $\gamma_1 + \gamma_2 = 1$, i.e., one of $\gamma_p = 0$ and other is equal to one. Consider that

$$uv^{\alpha_1+\alpha_2+1} wx^{\alpha_3+\alpha_4} y, uv^{\alpha_1+\alpha_2+1+\beta} wx^{\alpha_3+\alpha_4+\beta} y \in L.$$

Further, if $x = x_1x_2$ where $x_1, x_2 \in \Sigma^*$ and $|x_1| = j_2'$, then

$$uv^{\alpha_1+\alpha_2+1+\beta}wx^{\alpha_3+\alpha_4+\beta}y \in z_1 \cdot z_2 \cdot z_3 \sqcup\!\sqcup_t v^\beta (x_2x_1)^\beta$$

where

$$\begin{aligned} z_1 &= uv^{\alpha_1+\gamma_1}, \\ z_2 &= v^{\alpha_2+\gamma_2}wx^{\alpha_3}x_1, \\ z_3 &= x_2x^{\alpha_4-1}y, \\ t &= 0^{i+j_1(\alpha_1+\gamma_1)}1^{j_1\beta}0^{k+j_2'+j_1(\alpha_2+\gamma_2)+j_2\alpha_3}1^{j_2\beta}0^{\ell-j_2'+j_2\alpha_4}. \end{aligned}$$

By (6.17), $t \in T$. Note also that

$$z_1z_2z_3 = uv^{\alpha_1+\alpha_2+1}wx^{\alpha_3+\alpha_4}y \in L.$$

As $vx \neq \epsilon$ and $\beta > 0$, $v^\beta(x_2x_1)^\beta \neq \epsilon$. Thus, $L \notin \mathcal{P}_T$. ■

Note that if $T \supseteq 0^*1^*0^*1^*0^*$ then T satisfies the conditions of Lemma 6.9.4. This instance of our result is also a corollary of a result due to Thierrin and Yu [193, Prop. 3.3(2)].

6.9.3 Finiteness of T-codes

We now turn to the question of the existence of arbitrary infinite languages in a class of T-codes. We first show that if T is bounded, then there is an infinite T-code.

Theorem 6.9.5 *Let $T \subseteq \{0,1\}^*$ be a bounded set of trajectories. Then for all Σ with $|\Sigma| > 1$, $\mathcal{P}_T(\Sigma)$ contains an infinite language.*

Proof. Let $T \subseteq \{0,1\}^*$ be a bounded language. Then there exist $k \geq 0$ and $w_1, w_2, \ldots, w_k \in \{0,1\}^*$ such that $T \subseteq w_1^*w_2^* \cdots w_k^*$. By Lemma 6.3.2, if we can establish that there is an infinite T'-code, where $T' = w_1^* \cdots w_k^*$, the result will follow. Thus, without loss of generality, we let $T = w_1^*w_2^* \cdots w_k^*$.

If $w_1 = w_2 = \cdots = w_k = \epsilon$, then $T = \{\epsilon\}$, and thus $\mathcal{P}_T(\Sigma) = 2^{\Sigma^+} - \{\emptyset\}$, which clearly contains an infinite language.

Otherwise, there exists i_0 with $1 \leq i_0 \leq k$ such that $w_{i_0} \neq \epsilon$. For all $1 \leq i \leq k$, let $\alpha_i = |w_i|$. Let $a, b \in \Sigma$ be distinct letters, and define $L_T \subseteq \{a, b\}^+$ by

$$L_T = \{(\prod_{i=1}^{k} a^m b^{\alpha_i}) a^m \ : \ m \geq 0\}.$$

We have that $L_T \subseteq \{a, b\}^+$ as $\alpha_{i_0} \neq 0$. We claim $L_T \in \mathcal{P}_T(\Sigma)$. Assume not. Then there exist $m_1, m_2 \in \mathbb{N}$ with $m_1 > m_2$, $t \in T$ and $z \in \Sigma^+$ such that

$$(\prod_{i=1}^{k} a^{m_1} b^{\alpha_i}) a^{m_1} \in (\prod_{i=1}^{k} a^{m_2} b^{\alpha_i}) a^{m_2} \sqcup_t z.$$

Thus, we have that $z = a^{(k+1)(m_1 - m_2)}$. Further, let $t_i \in \{0, 1\}^*$ for $1 \leq i \leq k+1$ be defined so that

$$t = (\prod_{i=1}^{k} t_i 0^{\alpha_i}) t_{k+1},$$

where $|t_i|_0 = m_2$ and $|t_i|_1 = m_1 - m_2$ for all $1 \leq i \leq k+1$. As $t \in T$, there exist $j_i \in \mathbb{N}$, $1 \leq i \leq k$, such that $t = \prod_{i=1}^{k} w_i^{j_i}$. Thus, we have that

$$\sum_{i=1}^{k} \alpha_i j_i = (\sum_{i=1}^{k} |t_i| + \alpha_i) + |t_{k+1}|,$$

and so

$$\sum_{i=1}^{k} \alpha_i j_i \geq \sum_{i=1}^{k} |t_i| + \alpha_i.$$

Let ℓ with $1 \leq \ell \leq k$ be the minimal index such that

$$\sum_{i=1}^{\ell} \alpha_i j_i \geq \sum_{i=1}^{\ell} |t_i| + \alpha_i. \tag{6.18}$$

Note that $j_\ell > 0$, since if $j_\ell = 0$, then $\ell - 1$ satisfies (6.18) as well, contrary to our choice of ℓ (if $\ell = 1$ and $j_1 = 0$, then $|t_1| = 0$, which is a contradiction to $|t_1| = m_1 > 0$).

Let $u_1 = \prod_{i=1}^{\ell-1} t_i 0^{\alpha_i}$, $u_2 = (\prod_{i=\ell+1}^{k} t_i 0^{\alpha_i}) t_{k+1}$, $s_1 = \prod_{i=1}^{\ell-1} w_i^{j_i}$ and $s_2 = \prod_{i=\ell+1}^{k} w_i^{j_i}$. Thus, we have that

$$u_1 t_\ell 0^{\alpha_\ell} u_2 = s_1 w_\ell^{j_\ell} s_2$$

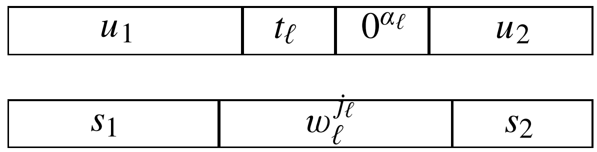

Figure 6.1: Two factorizations of t.

with $|u_1| \geq |s_1|$ and $|u_1| + |t_\ell| + \alpha_\ell \leq |s_1| + \alpha_\ell \cdot j_\ell$. The situation is summarized in Fig. 6.1. Thus, we have that $w_\ell^{j_\ell}$ contains a block of zeroes of length α_ℓ. As $|w_\ell| = \alpha_\ell$ and $j_\ell \neq 0$, this implies that $w_\ell = 0^{\alpha_\ell}$. But then as t_ℓ is a factor of $w_\ell^{j_\ell}$, we also have that $t_\ell \in 0^*$. Thus, $|t_\ell|_1 = 0$, and $m_1 = m_2$, a contradiction. ■

Further, there exist uncountably many unbounded trajectories T such that $\mathcal{P}_T$ contains infinite–even infinite *regular*–languages. Infinitely many of these are unbounded regular sets of trajectories.

Theorem 6.9.6 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that there exists $n \geq 0$ such that $T \subseteq 0^{\leq n}1(0+1)^*$. Then for all Σ with $|\Sigma| > 1$, $\mathcal{P}_T(\Sigma)$ contains an infinite regular language.*

Proof. Let $n \geq 0$ and $T^{(n)} = 0^{\leq n}1(0+1)^*$. By Lemma 6.3.2, it suffices to prove that $\mathcal{P}_{T^{(n)}}(\Sigma)$ contains an infinite regular language. Let $a, b \in \Sigma$ be distinct letters. Consider the regular language $R_n = a^{n+1}b^*$. Assume that $R_n \notin \mathcal{P}_{T^{(n)}}(\Sigma)$. Thus, there exist $i \geq 0$, $t_0 \in T^{(n)}$ and $z \in \{a,b\}^+$ such that $(a^{n+1}b^i \sqcup_{t_0} z) \cap R_n \neq \emptyset$. Let $t_0 = 0^m 1 t_2$ for some $n \geq m \geq 0$ and $t_2 \in \{0,1\}^*$. Consider that

$$a^{n+1}b^i \sqcup_{t_0} z = a^m z_1 (a^{n+1-m} b^i \sqcup_{t_2} z_2)$$

where $z = z_1 z_2$ and $z_1 \in \{a, b\}$.

By assumption, $a^m z_1 (a^{n+1-m} b^i \sqcup_{t_2} z_2) \cap R_n \neq \emptyset$, so that $z_1 = a$. But now,

$$(a^{n+1-m} b^i \sqcup_{t_2} z_2) \cap a^{n-m} b^* \neq \emptyset,$$

which is clearly impossible, since $|x|_a \geq n+1-m$ for all $x \in a^{n+1-m} b^i \sqcup_{t_2} z_2$. ■

The following corollary holds by Lemma 6.7.6.

Corollary 6.9.7 *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that there exists $n \geq 0$ such that $T \subseteq (0+1)^* 10^{\leq n}$. Then for all Σ with $|\Sigma| > 1$, $\mathcal{P}_T(\Sigma)$ contains an infinite regular language.*

We now turn to defining sets T of trajectories such that all T-codes are finite. The following proof is generalized from the case $H = (0+1)^*$ found in, e.g., Lothaire [140] or Conway [28, pp. 63–64].

Lemma 6.9.8 *Let $n, m \geq 1$ be such that $m \mid n$. Let $T_{n,m} = (0^n + 1^m)^* 0^{\leq n-1}$. Then $T_{n,m} \in \mathfrak{F}_H$.*

Proof. In what follows, let $\omega = \omega_{T_{n,m}}$. Assume that there exists an infinite $T_{n,m}$-code. Then there exists an infinite sequence $\{x_i\}_{i \geq 1}$ which is ω-free, i.e., $i < j$ implies $x_i \ \omega \ x_j$ does not hold. As $T_{n,m} \supseteq 0^*$, ω is reflexive and we have that $x_i \neq x_j$ for all $i > j \geq 1$.

We now choose (using the axiom of choice) a minimal infinite ω-free sequence as follows: let y_1 be the shortest word which begins an infinite ω-free sequence. Let y_2 be the shortest word such that y_1, y_2 begins an infinite ω-free sequence. We continue in this way. Let $\{y_i\}_{i \geq 1}$ be the resulting sequence. Clearly, $\{y_i\}_{i \geq 1}$ is an infinite ω-free sequence.

As ω is reflexive, $y_i \neq y_j$ for all $i > j \geq 1$. Therefore, $|y_i| \leq n$ for only finitely many $i \in \mathbb{N}$. Furthermore, since there are only finitely many words of length n, there exist $y \in \Sigma^n$ and $\{i_j\}_{j \geq 1} \subseteq \mathbb{N}$ such that y is a prefix of y_{i_j} for all $j \geq 1$. In particular, for all $j \geq 1$, let $u_j \in \Sigma^*$ be the word such that $y_{i_j} = y u_j$. Consider the sequence

$$Y = \{y_1, y_2, y_3, \cdots, y_{i_1 - 1}, u_1, u_2, \cdots\}.$$

Clearly, as $n \geq 1$, $|u_1| < |y_{i_1}|$. Thus, Y is an infinite sequence which comes before $\{y_i\}_{i \geq 1}$ in our ordering of infinite ω-free sequences, and so two words in Y must be comparable under ω. By assumption, $y_{j_1} \not\omega \ y_{j_2}$ for all $1 \leq j_1 < j_2 \leq i_1 - 1$. Thus, there are two remaining cases:

(i) there exist $1 \leq j \leq i_1 - 1$ and $k \geq 1$ such that $y_j \omega u_k$. Thus, let $t \in T_{n,m}$ and $\alpha \in \Sigma^*$ be chosen so that $u_k \in y_j \sqcup\!\sqcup_t \alpha$. Consider $t' = 1^n t \in T_{n,m}$. Then $y_{i_k} = y u_k \in y(y_j \sqcup\!\sqcup_t \alpha) = y_j \sqcup\!\sqcup_{t'} y\alpha$. Therefore, $y_j \ \omega \ y_{i_k}$. As $j \leq i_1 - 1 < i_k$, this is a contradiction.

(ii) there exist $k > \ell \geq 1$ such that $u_\ell \, \omega \, u_k$. Let $\alpha \in \Sigma^*$ and $t \in T_{n,m}$ be such that $u_k \in u_\ell \sqcup\!\sqcup_t \alpha$. Consider $t' = 0^n t \in T_{n,m}$. Then $y_{i_k} = yu_k \in y(u_\ell \sqcup\!\sqcup_t \alpha) = yu_\ell \sqcup\!\sqcup_{t'} \alpha = y_{i_\ell} \sqcup\!\sqcup_{t'} \alpha$. Thus $y_{i_\ell} \, \omega \, y_{i_k}$. As $\ell < k$, this is a contradiction.

We have arrived at a contradiction. ■

As another class of examples, Ehrenfeucht *et al.* [47, p. 317] note that $\{1^n, 0\}^* \in \mathfrak{F}_H$ for all $n \geq 1$ (their other results, though elegant and interesting, do not otherwise seem to be applicable to our situation).

Note that $T_{1,1} = \{0,1\}^*$. Let $T_n = T_{n,n}$. For all $1 \leq i < j$, $\mathcal{P}_{T_i} \neq \mathcal{P}_{T_j}$, as $0^i 1^i \in T_i - T_j$. Thus, by Lemma 6.3.6, the classes of T_i- and T_j-codes are distinct.

Corollary 6.9.9 *There are infinitely many $T \subseteq \{0,1\}^*$ which define distinct classes $\mathcal{P}_T$ satisfying $\mathcal{P}_T \subseteq$ FIN.*

Further, the following is immediate:

Corollary 6.9.10 *Let $T \subseteq \{0,1\}^*$ be such that $T_n \subseteq T$ for some $n \geq 1$. Then $\mathcal{P}_T \subseteq$ FIN.*

Ilie [73, Sect. 7.7] also gives a class of partial orders which we may phrase in terms of sets of trajectories. In particular, define the set of functions

$$\mathcal{G} = \{g : \mathbb{N} \to \mathbb{N} \ : \ g(0) = 0 \text{ and } 1 \leq g(n) \leq n \text{ for all } n \geq 1\}.$$

Then for all $g \in \mathcal{G}$, we define

$$T_g = \{1^* \prod_{k=1}^{m} (0^{i_k} 1^*) \ : \ i_k \geq 0 \, \forall 1 \leq k \leq m; \ m = g(\sum_{k=1}^{m} i_k)\}.$$

We denote the *upper limit* of a sequence $\{s_n\}_{n \geq 1}$ by $\overline{\lim}_{n\to\infty} s_n$. We have the following result [73, Thm. 7.7.8]:

Theorem 6.9.11 *Let $g \in \mathcal{G}$. Then $T_g \in \mathfrak{F}_H \iff \overline{\lim}_{n\to\infty} \frac{n}{g(n)} < \infty$.*

6.9.4 Decidability and Finiteness Conditions

We now consider decidability of membership in $\mathcal{P}_T$ if T satisfies the conditions of the previous sections. We have the following positive decidability results:

Theorem 6.9.12 *Let T be recursive. If $T \in \mathfrak{F}_R$ (resp., $T \in \mathfrak{F}_C$, $T \in \mathfrak{F}_H$) then given a regular (resp., context-free, context-free) language L, it is decidable whether $L \in \mathcal{P}_T$.*

Proof. We establish the result for $T \in \mathfrak{F}_C$. The case $T \in \mathfrak{F}_H$ is an instance of this case and the case $T \in \mathfrak{F}_R$ is very similar. Let $T \in$ REC and $T \in \mathfrak{F}_C$. Let $L \in$ CF. We first check if L is infinite. If it is, then certainly $L \notin \mathcal{P}_T$, so we answer no.

If L is finite, then we can effectively find a list of all words in L (consider putting L in Chomsky Normal Form (CNF); see Hopcroft and Ullman [68] for an introduction to CNF). Let $F = L$, where F is some effectively given finite set. Then by Lemma 6.3.10, we can decide whether $L = F \in \mathcal{P}_T$. ■

One might hope for an undecidability result of the following type, which would complement Theorem 6.3.9: for a fixed $T \in$ REG (perhaps with some reasonable assumption, e.g., completeness), then it is undecidable, given a CFL L, whether $L \in \mathcal{P}_T$. Theorem 6.9.12 shows us that we cannot hope for a simple such result, since we need to restrict ourselves to those T which do not lie in $\mathfrak{F}_C$ in this case.

6.9.5 Up and Down Sets

Let $L \subseteq \Sigma^*$ and $T \subseteq \{0, 1\}^*$ Define $\text{DOWN}_T(L)$, $\text{UP}_T(L)$ as

$$\text{DOWN}_T(L) = L \rightsquigarrow_{\tau(T)} \Sigma^*;$$

$$\text{UP}_T(L) = L \sqcup\!\sqcup_T \Sigma^*.$$

Our notation roughly follows Harju and Ilie [59], where $\text{DOWN}_T(L)$ is denoted $\text{DOWN}_{\omega_T}(L)$ and $\text{UP}_T(L)$ is denoted $\text{DOWN}_{\omega_T^{-1}}(L)$.

Our aim in this section is, given T, to characterize the complexity $\textsc{up}_T(L)$ and $\textsc{down}_T(L)$ for arbitrary L. We will have a particular interest in those $T \in \mathfrak{F}_H$ which are partial orders. Let $\mathfrak{F}_H^{(po)}$ denote the class of all trajectories $T \in \mathfrak{F}_H$ which are partial orders.

Haines [58] observed that for $T = (0+1)^*$, $\textsc{up}_T(L)$ and $\textsc{down}_T(L)$ are regular languages for all L. There is an elegant generalization of Haines' result due to Harju and Ilie [59]: If we restrict our attention to those $T \in \mathfrak{F}_H$ which are compatible, then $\textsc{up}_T(L)$ and $\textsc{down}_T(L)$ are still regular languages for all languages L. We recall this in the following result, which is a specific case of a result due to Harju and Ilie [59, Thm. 6.3]:[2]

Theorem 6.9.13 *Let $T \in \mathfrak{F}_H$ be compatible. Let $L \subseteq \Sigma^*$ be a language. Then $\textsc{up}_T(L)$, $\textsc{down}_T(L)$ are regular languages.*

The following corollary is an interesting consequence:

Corollary 6.9.14 *Let $T \in \mathfrak{F}_H$ satisfy $0^* \subseteq T^*$. Let $L \subseteq \Sigma^*$ be a language. Then the languages $\textsc{up}_{T^*}(L)$, $\textsc{down}_{T^*}(L)$ are regular.*

Proof. If $0^* \subseteq T^*$ then T^* is clearly compatible by Corollary 6.4.14. Further, as $T \subseteq T^*$, we have $T^* \in \mathfrak{F}_H$. The result now follows by Theorem 6.9.13. ∎

We now consider arbitrary $T \in \mathfrak{F}_H^{(po)}$ and seek to characterize the complexity of $\textsc{up}_T(L)$ and $\textsc{down}_T(L)$. By the same proofs as given for $H = (0+1)^*$ (see, e.g., Harrison [62, Sect. 6.6]), we have the following results:

Lemma 6.9.15 *Let $T \subseteq \mathfrak{F}_H^{(po)}$. Let $L \subseteq \Sigma^*$. Then*

(a) there exists a finite language $F \subseteq \Sigma^*$ such that $\textsc{up}_T(L) = \textsc{up}_T(F)$.

(b) there exists a finite language $G \subseteq \Sigma^*$ such that $\textsc{down}_T(L) = \overline{\textsc{up}_T(G)}$.

We now characterize the complexity of $\textsc{up}_T(L)$ and $\textsc{down}_T(L)$ for all L, based on the complexity of T:

[2] Note that what Harju and Ilie call monotone, we call compatible.

Theorem 6.9.16 *Let $\mathcal{C}$ be a cone. Let $T \in \mathfrak{F}_H^{(po)}$ be an element of $\mathcal{C}$. Then for all $L \subseteq \Sigma^*$, $\text{UP}_T(L) \in \mathcal{C}$ and $\text{DOWN}_T(L) \in$ co-$\mathcal{C}$.*

Proof. Let $L \subseteq \Sigma^*$. Then there exists $F \subseteq \Sigma^*$ such that $\text{UP}_T(L) = \text{UP}_T(F) = F \sqcup\!\sqcup_T \Sigma^*$. By the closure properties of cones under $\sqcup\!\sqcup_T$, $\text{UP}_T(L) \in \mathcal{C}$. A similar proof shows that $\text{DOWN}_T(L) \in$ co-$\mathcal{C}$. ■

6.9.6 T-Convexity Revisited

We now turn to the complexity of T-convex languages:

Theorem 6.9.17 *Let $\mathcal{C}$ be a cone. Let $T \in \mathfrak{F}_H^{(po)}$ be an element of $\mathcal{C}$. Then every T-convex language is an element of $\mathcal{C} \wedge$ co-$\mathcal{C}$.*

Proof. Let $T \in \mathfrak{F}_H^{(po)}$. As T is a partial order, it is reflexive. Thus, if L is a T-convex language, we have that $L = \text{UP}_T(L) \cap \text{DOWN}_T(L)$ by Corollary 6.6.2. Thus, by Theorem 6.9.16, the result follows. ■

The following corollary is immediate, based on the closure properties of the recursive and regular languages:

Corollary 6.9.18 *Let $T \in \text{REG}$ (resp., REC) be such that $T \in \mathfrak{F}_H^{(po)}$. If L is a T-convex language, then $L \in \text{REG}$ (resp., REC).*

Corollary 6.9.18 was known for the case of $H = (0+1)^*$ and $L \in \text{REG}$, see Thierrin [192, Cor. to Prop. 3]. Further, we can also establish the following result:

Theorem 6.9.19 *Let $T \in \mathfrak{F}_H$ be compatible. Then every T-convex language is regular.*

Consider the sets $E_n = \{0, 1^n\}^*$. As noted by Ehrenfeucht *et al.* [47], $E_n \in \mathfrak{F}_H$. As $E_n = E_n^*$ and $0^* \subseteq E_n$, E_n is compatible. Thus, we have that every E_n-convex language is regular.

6.10 Conclusions

We have introduced the notion of a T-code, and examined its properties. Many results which are known in the literature are specific instances of general results on T-codes. However, the notion of a T-code is not so general as to prevent interesting results from being obtained. We feel that the framework of T-codes is very suitable for further analysis of the general structure of the many classes of codes which it generalizes. Further research into this area should prove very useful.