# Chapter 8

# Iteration of Trajectory Operations

## 8.1 Introduction

Iterated concatenation, known as Kleene closure, is one of the defining operations of regular languages, and its properties are well known. As is commonly noted, "regular expressions without the star operator define only finite languages " [201, p. 77] (others, e.g., Salomaa [175] also express the same idea). There are many fundamental and deep results in formal language theory related to Kleene closure: we mention only the study of primitive words and star-height as examples. In this chapter, we examine iteration of trajectory-based operations, such as iterated (arbitrary) shuffle, which has been a topic for active research for the past 25 years [55, 85, 86, 87, 88, 93, 170, 182, 198].

We generalize the study of quotients and residuals with respect to an operation, which have been studied for particular operations by Câmpeanu *et al.* [21], Ito *et al.* [78, 79] and Kari and Thierrin [115]. We show that the smallest language containing $L$ and closed under shuffle along $T$ (or deletion along $T$) is the (positive) iteration closure of $L$ under $T$.

We also examine the concepts of *shuffle bases* and *extended shuffle bases*. These have been previously studied by Ito *et al.* [82], Ito *et al.* [78, 79], Ito and Silva [80] and Hsiao *et al.* [69]. These notions are related to the concept of $T$-codes introduced in Chapter 6.

Some of the work in this chapter has previously appeared in the more general setting of *word* *operations*, as studied by Hsiao *et al.* [69]. However, we present the results below for several important reasons. First, the framework on shuffle on trajectories and deletion along trajectories yields closure properties which do not necessarily hold in the more general setting of word operations. Further, we have presented our results with slightly modified definitions which we feel are more natural. These modified definitions allow us to drop certain assumptions which were necessary in the setting of word operations, and also allow us to make interesting conclusions to the classes of $T$-codes which were not done in the more general setting.

## 8.2 Definitions

We first define the iterated shuffle operations relative to a given set $T$ of trajectories. Let $T \subseteq \{0,1\}^*$ be a set of trajectories. Then, for all languages $L \subseteq \Sigma^*$ and all $i \geq 0$, we define $(\sqcup\!\sqcup_T)^i(L)$ as follows:

$$\begin{aligned}(\sqcup\!\sqcup_T)^0(L) &= \{\epsilon\} \\ (\sqcup\!\sqcup_T)^1(L) &= L \\ (\sqcup\!\sqcup_T)^{i+1}(L) &= \big((\sqcup\!\sqcup_T)^i(L) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^i(L)\big) \cup (\sqcup\!\sqcup_T)^i(L) \quad \forall i \geq 1. \end{aligned} \tag{8.1}$$

Note that we do not require that $T$ defines an associative operation. Further, we define $(\sqcup\!\sqcup_T)^*(L)$ and $(\sqcup\!\sqcup_T)^+(L)$ as

$$\begin{aligned}(\sqcup\!\sqcup_T)^*(L) &= \bigcup_{i\geq 0}(\sqcup\!\sqcup_T)^i(L); \\ (\sqcup\!\sqcup_T)^+(L) &= \bigcup_{i\geq 1}(\sqcup\!\sqcup_T)^i(L).\end{aligned}$$

Similarly, we define iterated deletion along a set $T$ of trajectories. Let $T \subseteq \{i,d\}^*$ be a set of trajectories. Then, for all $L \subseteq \Sigma^*$ and all $i \geq 0$, we define $(\rightsquigarrow_T)^i(L)$ as follows:

$$\begin{aligned}(\rightsquigarrow_T)^0(L) &= \{\epsilon\}; \\ (\rightsquigarrow_T)^1(L) &= L; \\ (\rightsquigarrow_T)^{i+1}(L) &= \big((\rightsquigarrow_T)^i(L) \rightsquigarrow_T (\rightsquigarrow_T)^i(L)\big) \cup (\rightsquigarrow_T)^i(L) \quad \forall i \geq 1. \end{aligned} \tag{8.2}$$

We again do not require that $T$ defines an associative operation. Further, we define $(\rightsquigarrow_T)^*(L)$ and $(\rightsquigarrow_T)^+(L)$ as

$$\begin{aligned}(\rightsquigarrow_T)^*(L) &= \bigcup_{i\geq 0}(\rightsquigarrow_T)^i(L);\\ (\rightsquigarrow_T)^+(L) &= \bigcup_{i\geq 1}(\rightsquigarrow_T)^i(L).\end{aligned}$$

We also require an auxiliary operation $L_1[\rightsquigarrow_T]^i L_2$ which is defined recursively for all $i \geq 0$ as follows:

$$\begin{aligned}L_1[\rightsquigarrow_T]^0 L_2 &= L_1\\ L_1[\rightsquigarrow_T]^{i+1} L_2 &= (L_1[\rightsquigarrow_T]^i L_2) \rightsquigarrow_T L_2 \quad \forall i \geq 1.\end{aligned}$$

We then set

$$L_1[\rightsquigarrow_T]^* L_2 = \bigcup_{i\geq 0} L_1[\rightsquigarrow_T]^i L_2.$$

## 8.3 Iterated Shuffle on Trajectories

We begin our investigation with iterated shuffle on trajectories. We require some preliminary discussion regarding our definition and an alternate definition. Then we discuss some examples of iterated shuffle on trajectories before beginning our examination of the operation.

### 8.3.1 Left-Associativity and a Simplified Definition

Let $T \subseteq \{0,1\}^*$. We say that $T$ is *left-associative* if, for all $\alpha, \beta, \gamma \in \Sigma^*$,

$$\alpha \sqcup\!\sqcup_T (\beta \sqcup\!\sqcup_T \gamma) \subseteq (\alpha \sqcup\!\sqcup_T \beta) \sqcup\!\sqcup_T \gamma.$$

Note that associativity implies left-associativity. Further, we can verify that $T = 0^*1^*0^*$ (insertion) is left-associative but not associative. Initial literal shuffle, given by $T = (01)^*(0^* + 1^*)$, is not left-associative. Left-associativity is also called left-inclusiveness [69]. Several of the results obtained by Hsiao *et al.* [69] are similar to our results, but must include the condition of left-associativity, due to a slightly less general definition of iterated operations, given by the recurrence

$$\begin{aligned}(\sqcup_T)^0_X(L) &= \{\epsilon\}\\ (\sqcup_T)^1_X(L) &= L\\ (\sqcup_T)^{i+1}_X(L) &= (\sqcup_T)^i_X(L) \sqcup_T L \quad \forall i \geq 1 \end{aligned} \tag{8.3}$$

instead of (8.1). Definitions (8.1 and (8.3) agree when the operation is left-associative. For example, our Theorem 8.6.5 in Section 8.6.2 below is given by Hsiao *et al.* [69] for left-associative word operations.

We now give a characterization of left-associativity. It is a weakening of the corresponding result of Mateescu *et al.* [147] on associativity. Let $D = \{x, y, z\}$. Then let $\tau, \sigma, \varphi, \psi : D^* \to \{0, 1\}^*$ be the morphisms given by

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

Then the following result follows by the same proof as in Mateescu *et al.* [147, Prop. 4.7]:

**Theorem 8.3.1** *Let $T \subseteq \{0, 1\}^*$. Then $T$ is left-associative if and only if*

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

*Thus, if $T$ is regular, it is decidable if $T$ is left-associative.*

Let $(\sqcup_T)^*_X, (\sqcup_T)^+_X$ be the iterated versions of $\sqcup_T$ defined by (8.3) instead of (8.1), i.e.,

$$(\sqcup_T)^*_X(L) = \bigcup_{i\geq 0} (\sqcup_T)^i_X(L); \tag{8.4}$$

$$(\sqcup_T)^+_X(L) = \bigcup_{i\geq 1} (\sqcup_T)^i_X(L). \tag{8.5}$$

(8.6)

Again, we note that it is not hard to establish that $(\sqcup_T)^*_X(L) = (\sqcup_T)^*(L)$ for all $L$ if $T$ is left-associative.

### 8.3.2 Some examples

We begin by noting the most well-studied iteration operation, that of Kleene closure. If $T = 0^*1^*$, then $\sqcup\!\sqcup_T$ defines the concatenation operation. Note that $T = 0^*1^*$ is associative. Thus

$$(\cdot)^*(L) = L^* = \{w_1 w_2 w_3 \cdots w_n \ : \ n \geq 0, w_i \in L\}.$$

We note that if $L$ is regular, then $L^*$ is regular.

If $T = (0+1)^*$, we get the operation of shuffle-closure, which has been well-studied in the literature (see Gischer [55], Jędrzejowicz [87, 88, 89, 91, 92], Kari and Thierrin [118], Ito *et al.* [79] as well as much work in software specification [6, 83, 182, 170]). Let us denote this case by $(\sqcup\!\sqcup)^*(L)$. We note that $(\sqcup\!\sqcup)^*(L)$ does not preserve regularity, even if $L$ is a singleton set, as

$$(\sqcup\!\sqcup)^*(\{ab\}) \cap a^*b^* = \{a^n b^n \ : \ n \geq 0\}.$$

We also note that the CFLs are not closed under $(\sqcup\!\sqcup)^*$, in fact, there exist a finite set $F$ such that $(\sqcup\!\sqcup)^*(F)$ is not a CFL. Let $L = \{abc, acb, bac, bca, cab, cba\}$. Then we can see that

$$(\sqcup\!\sqcup)^*(L) = \{w \in \{a,b,c\}^* \ : \ |w|_a = |w|_b = |w|_c\}.$$

Using a grammar-based argument, Gischer [55] proved that the closure of the regular languages under $(\sqcup\!\sqcup)^*$ is a proper subset of the CSLs. For an LBA-based proof of inclusion, see Jędrzejowicz [87]. We note that with a simple extension of the result of Jędrzejowicz, we can show that $(\sqcup\!\sqcup_T)^*(L) \in$ CS for all $T, L \in$ CS with $T$ left-associative.

If $T = 0^*1^*0^*$, we get the insertion operation, $\leftarrow$. Iterated insertion has been studied by Ito *et al.* [78], Kari and Thierrin [118] and Holzer and Lange [67]. Again, we note that $(\leftarrow)^*(\{ab\}) \cap a^*b^* = \{a^n b^n \ : \ n \geq 0\}$. Thus, iterated insertion of a singleton can result in non-regular sets.

If $T = 0^*1^* + 1^*0^*$, $\sqcup\!\sqcup_T$ is the bi-catenation operation (see Shyr and Yu [187] or Hsiao *et al.* [69]). Note that $L_1 \sqcup\!\sqcup_T L_2 = L_1L_2 + L_2L_1$. Thus,

$$(\sqcup\!\sqcup_T)^2(L) = L^2 + L^2 = L^2$$

and $(\sqcup\!\sqcup_T)^*(L) = L^*$. Thus, iterated bi-catenation preserves regularity.

### 8.3.3 Iteration and Density

We now turn to examining the relation of the density of a set of trajectories to the closure properties of its iteration operation. Recall that the density of a language was defined in Section 4.3. We begin by noting that preserving regularity is incomparable with density of $T$. We note that $T = 0^*1^*$ has density $O(n)$, and that the associated operation (Kleene closure) preserves regularity. However, there exist constant density sets of trajectories whose iteration closure does not preserve regularity, even when applied to finite sets:

**Lemma 8.3.2** *Let $T = 10^*1$. Then $p_T(n) = 1$ but $(\sqcup\!\sqcup_T)^*(\{ab\}) = \{a^n b^n \ : \ n \geq 0\}$.*

We now show that the set of trajectories in Lemma 8.3.2 can be considered the simplest constant-density regular set of trajectories which does not preserve regularity, when considering $(\sqcup\!\sqcup_T)^*_X$. In particular, we show that if $T$ has constant density, then $(\sqcup\!\sqcup_T)^*_X$ preserves finiteness unless $T \supseteq u(0^c)^*v$ for some $u, v \in \{0,1\}^*$ and $c \geq 1$.

**Lemma 8.3.3** *Let $T = \cup_{i=1}^k u_i v_i^* w_i$ for $u_i, v_i, w_i \in \{0,1\}^*$ and $|v_i|_1 > 0$ for all $1 \leq i \leq k$. Then for all finite languages $L$, $(\sqcup\!\sqcup_T)^*_X(L)$ is finite.*

**Proof.** For all $1 \leq i \leq k$, let $\beta_i = |u_i w_i|_1$ and $\eta_i = |v_i|_1 > 0$. Let $\beta = \max_{1 \leq i \leq k}\{\beta_i\}$ and $\eta = \max_{1 \leq i \leq k}\{\eta_i\}$.

Let $1 \leq i \leq k$. If, for all $x \in L$, there is no $\ell \geq 0$ such that $|x| = \beta_i + \ell\eta_i$, then for all $X \subseteq \Sigma^*$, $X \sqcup\!\sqcup_{u_i v_i^* w_i} L = \emptyset$. Thus, without loss of generality, we can assume that for each $1 \leq i \leq k$, there is some $x \in L$ and $\ell \geq 0$ such that $|x| = \beta_i + \ell\eta_i$, otherwise, we can replace $T$ with

$$T' = \bigcup_{\substack{1 \leq j \leq k \\ j \neq i}} u_j v_j^* w_j.$$

For all $1 \leq i \leq k$, let $\ell_i = \max\{\ell \geq 0 \ : \ \exists u \in L \text{ such that } |u| = \beta_i + \ell\eta_i\}$. Since $L$ is finite, $\ell_i$ exists. Let $\lambda = \max_{1 \leq i \leq k}\{\ell_i\}$.

Let $x \in (\sqcup\!\sqcup_T)^*_X(L)$. We show that the length of $x$ is bounded above by $\beta + \lambda\eta$. As $x \in (\sqcup\!\sqcup_T)^*_X(L)$, either $x = \epsilon$, or there is some $y \in (\sqcup\!\sqcup_T)^*_X(L)$ and $z \in L$ such that $x \in y \sqcup\!\sqcup_T z$. Let $1 \le i \le k$ and $s \ge 0$ be such that $x \in y \sqcup\!\sqcup_t z$ where $t = u_i v_i^s w_i$.

As $\eta_i \neq 0$, $|t|_1 = \beta_i + s\eta_i > \beta_i + \ell_i\eta_i$. As $|z| = |t|_1$, we have that $s \le \ell_i$ by choice of $\ell_i$. Now $|x| = |t| = \beta_i + s\eta_i \le \beta_i + \ell_i\eta_i \le \beta + \lambda\eta$. ■

Consider the following particular case of a result due to Szilard *et al.* [190]:

**Lemma 8.3.4** *Let $L \subseteq \Sigma^*$ be a regular language such that $p_L(n) \in O(1)$. Then $L$ is a finite union of terms of the form $uv^*w$ for words $u, v, w \in \Sigma^*$.*

Then, the following corollary is immediate.

**Corollary 8.3.5** *Let $T \subseteq \{0,1\}^*$ be a regular set of trajectories such that $p_T(n) \in O(1)$, and the closure of the finite languages under $(\sqcup\!\sqcup_T)^*_X$ contains an infinite language. Then $T \supseteq u(0^c)^*v$ for some $u, v \in \{0,1\}^*$ and $c \ge 1$.*

We note, however, that if we drop the condition that we use $(\sqcup\!\sqcup_T)^*_X$ instead of $(\sqcup\!\sqcup_T)^*$, the result no longer holds. Consider $T = (01)^*$. Then we have that $(\sqcup\!\sqcup_T)^*(\{ab\}) \supseteq \{a^n b^n \; : \; n \ge 0\}$.

We now turn to regular sets of trajectories whose iteration closure contains non-CF languages. Note that if $T = 10^*110^*$, $(\sqcup\!\sqcup_T)^*_X(\{abc\}) \cap a^*b^*c^* = \{a^n b^n c^n \; : \; n \ge 0\}$. Thus, there is a linear density regular set of trajectories whose iteration closure of singletons contains non-CF languages. However, we also have the following example: for $T = (01)^*$, $(\sqcup\!\sqcup_T)^*(\{abc\}) = \{a^{2^n} b^{2^n} c^{2^n} \; : \; n \ge 0\}$. We summarize the minimal known density of regular languages and regular sets of trajectories witnessing non-closure properties for iterated shuffle on trajectories in Table 8.1.

## 8.4 Iterated Deletion

We now consider iterated deletion operations. We first note that the finite languages are closed under iterated deletion:

| | $(\sqcup\!\sqcup_T)^*$ | | $(\sqcup\!\sqcup_T)^*_X$ | |
|---|---|---|---|---|
| | $p_T(n)$ | $p_L(n)$ | $p_T(n)$ | $p_L(n)$ |
| non-regular | $O(1)$ | $O(1)$ | $O(1)$ | $O(1)$ |
| non-CF | $O(1)$ | $O(1)$ | $O(n)$ | $O(1)$ |

Figure 8.1: Summary of minimum-density regular languages and regular sets of trajectories demonstrating non-closure properties for iterated shuffle on trajectories.

**Lemma 8.4.1** *Let $L \subseteq \Sigma^{\leq m}$ for some $m \geq 0$. Then for all $T \subseteq \{i, d\}^*$, $(\rightsquigarrow_T)^*(L) \subseteq \Sigma^{\leq m}$.*

Second, we show that an alternate definition will suffice for some operations we will consider here. This alternate definition will somewhat simplify the results in this section. Call a set of trajectories $T \subseteq \{i, d\}^*$ *del-left-preserving* if $T \supseteq i^*$. Consider the following definitions:

$$(\rightsquigarrow_T)^0_X(L) = \{\epsilon\}$$

$$(\rightsquigarrow_T)^1_X(L) = L$$

$$(\rightsquigarrow_T)^{i+1}_X(L) = (\rightsquigarrow_T)^i_X(L) \rightsquigarrow_T ((\rightsquigarrow_T)^i_X(L) \cup \{\epsilon\}) \forall i \geq 1 \tag{8.7}$$

We also define $(\rightsquigarrow_T)^*_X$ and $(\rightsquigarrow_T)^+_X$:

$$(\rightsquigarrow_T)^*_X(L) = \bigcup_{i \geq 0} (\rightsquigarrow_T)^i_X(L); \tag{8.8}$$

$$(\rightsquigarrow_T)^+_X(L) = \bigcup_{i \geq 1} (\rightsquigarrow_T)^i_X(L). \tag{8.9}$$

The following result motivates the above definitions:

**Theorem 8.4.2** *Let $T \subseteq \{i, d\}^*$ be a del-left-preserving set of trajectories. Then for all $L \subseteq \Sigma^*$, $(\rightsquigarrow_T)^*(L) = (\rightsquigarrow_T)^*_X(L)$.*

**Proof.** The result is immediate on noting the following two identities, which are obvious from the definition of $\rightsquigarrow_T$:

$$X \rightsquigarrow_T (Y + Z) = X \rightsquigarrow_T Y + X \rightsquigarrow_T Z; \tag{8.10}$$

$$X \rightsquigarrow_T \{\epsilon\} = X \rightsquigarrow_{T \cap i^*} \{\epsilon\}. \tag{8.11}$$

Further, it is clear that $X \rightsquigarrow_{i^*} \{\epsilon\} = X$. ■

### 8.4.1 Iterated Scattered Deletion

In this section, we consider a problem of Ito *et al.* [79] on iterated scattered deletion[1]. Recall that if $T = (0+1)^*$, we denote $\rightsquigarrow_T$ by $\rightsquigarrow$. Ito *et al.* [79] asked whether the regular languages are closed under $(\rightsquigarrow)^+$. We show that they are not.

Let $k \geq 2$ be arbitrary, and let $\Sigma_k = \{\alpha_i, \beta_i, \gamma_i, \eta_i\}_{i=1}^k$. Then we define $L_k \subseteq \Sigma_k^*$ as

$$L_k = \prod_{i=1}^k (\alpha_i \beta_i)^* \prod_{i=1}^k (\gamma_i \eta_i)^* + \bigcup_{i=1}^k \beta_i \eta_i.$$

We claim that

$$(\rightsquigarrow)^+(L_k) \cap \prod_{i=1}^k \alpha_i^+ \prod_{i=1}^k \gamma_i^+ = \{\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_k^{i_k}\gamma_1^{i_1}\gamma_2^{i_2}\cdots\gamma_k^{i_k} \;:\; i_j \geq 1\}. \tag{8.12}$$

and that $(\rightsquigarrow)^+(L_k)$ cannot be expressed as the intersection of $k-1$ context-free languages.

We first establish (8.12). Let $(i_1, i_2, \cdots, i_k) \in \mathbb{N}^k$. Then note that

$$\prod_{j=1}^k \alpha_j^{i_j} \prod_{j=1}^k \gamma_j^{i_j} \in (\cdots(\prod_{j=1}^k (\alpha_j\beta_j)^{i_j} \prod_{j=1}^k (\gamma_j\eta_j)^{i_j})[\rightsquigarrow]^{i_1}\beta_1\eta_1)\cdots)[\rightsquigarrow]^{i_k}\beta_k\eta_k.$$

Intuitively, we delete matching pairs of $\beta_j$ and $\eta_j$ from the word $\theta = \prod_{j=1}^k (\alpha_j\beta_j)^{i_j} \prod_{j=1}^k (\gamma_j\eta_j)^{i_j}$, and leave only occurrences of $\alpha_j$ and $\gamma_j$, of which we must necessarily have equal numbers, by choice of our word $\theta$. This establishes the right-to-left inclusion of (8.12). We now show the reverse inclusion. First, note that if $\theta \in (\rightsquigarrow)^+(L_k)$, then we can write $\theta = x_1x_2\cdots x_k y_1 y_2 \cdots y_k$ where $x_i \in \{\alpha_i, \beta_i\}^*$ and $y_i \in \{\gamma_i, \eta_i\}^*$. To prove the left-to-right inclusion of (8.12), we will establish the following stronger claim:

**Claim 8.4.3** *Let $x_1x_2\cdots x_k y_1y_2\cdots y_k \in (\rightsquigarrow)^+(L_k)$ where $x_i \in \{\alpha_i, \beta_i\}^*$ and $y_i \in \{\gamma_i, \eta_i\}^*$ for all $1 \leq i \leq k$. Then for all $1 \leq i \leq k$, the following equalities hold:*

$$|x_i|_{\alpha_i} - |x_i|_{\beta_i} = |y_i|_{\gamma_i} - |y_i|_{\eta_i}. \tag{8.13}$$

[1]Note: Since this research appeared in Bull. EATCS [36], I have been informed that this problem has been previously solved; see Ito and Silva [80], where the authors show that there exists a regular language $R$ such that $(\rightsquigarrow)^+(R)$ is not a CFL. We note that the results here were found independently, and extend those of Ito and Silva.

**Proof.** Let $z = x_1 x_2 \cdots x_k y_1 y_2 \cdots y_k \in (\leadsto)^+(L_k)$. Then there exists some $i \geq 1$ such that $z \in (\leadsto)^i(L_k)$. The proof is by induction on $i$. For $i = 1$, $z \in L_k$. Thus, we see that either

(a) for all $1 \leq \ell \leq k$, $x_\ell = (\alpha_\ell \beta_\ell)^{j_\ell}$ for some $j_\ell \geq 0$ and $y_\ell = (\gamma_\ell \eta_\ell)^{j'_\ell}$ for some $j'_\ell \geq 0$, in which case $|x_\ell|_{\alpha_\ell} - |x_\ell|_{\beta_\ell} = 0 = |y_\ell|_{\gamma_\ell} - |y_\ell|_{\eta_\ell}$; or

(b) $z = \beta_\ell \eta_\ell$ for some $1 \leq \ell \leq k$. Thus, $|x_\ell|_{\alpha_\ell} - |x_\ell|_{\beta_\ell} = -1 = |y_\ell|_{\gamma_\ell} - |y_\ell|_{\eta_\ell}$ and $x_j = y_j = \epsilon$ for all $1 \leq j \leq k$ with $j \neq \ell$.

Thus, the result holds for $i = 1$.

Assume the claim holds for all natural numbers less than $i$. Let $z \in (\leadsto)^i(L_k)$. Then there exists some $\theta \in (\leadsto)^{i-1}(L_k)$ and $\zeta \in (\leadsto)^{i-1}(L_k) \cup \{\epsilon\}$ such that $z \in \theta \leadsto \zeta$. If $\zeta = \epsilon$, then $z = \theta$ and the result holds by induction. Thus, let

$$\theta = u_1 u_2 \cdots u_k v_1 v_2 \cdots v_k$$

$$\zeta = s_1 s_2 \cdots s_k t_1 t_2 \cdots t_k$$

with $u_\ell, s_\ell \in \{\alpha_\ell, \beta_\ell\}^*$ and $v_\ell, t_\ell \in \{\gamma_\ell, \eta_\ell\}^*$ for all $1 \leq \ell \leq k$. Then note that for all $1 \leq \ell \leq k$,

$$|x_\ell|_{\alpha_\ell} = |u_\ell|_{\alpha_\ell} - |s_\ell|_{\alpha_\ell};$$

$$|x_\ell|_{\beta_\ell} = |u_\ell|_{\beta_\ell} - |s_\ell|_{\beta_\ell};$$

$$|y_\ell|_{\gamma_\ell} = |v_\ell|_{\gamma_\ell} - |t_\ell|_{\gamma_\ell};$$

$$|y_\ell|_{\eta_\ell} = |v_\ell|_{\eta_\ell} - |t_\ell|_{\eta_\ell}.$$

Thus, by induction, we can easily establish that the desired equalities hold. ■

We now show that $(\leadsto)^+(L_k)$ cannot be expressed as the intersection of $k - 1$ context-free languages. Let $\mathrm{CF}_k$ be the class of languages which are expressible as the intersection of $k$ CFLs. The following lemma is obvious, since the CFLs are closed under intersection with regular languages, (for further closure properties of $\mathrm{CF}_k$, see, e.g., Latta and Wall [129]).

**Lemma 8.4.4** $\mathrm{CF}_k$ *is closed under intersection with regular languages.*

We will also require the following lemma:

**Lemma 8.4.5** *Let $L_1, L_2 \in \mathrm{CF}_k$ be such that there exist disjoint regular languages $R_1, R_2$ such that $L_i \subseteq R_i$ for $i = 1, 2$. Then $L_1 \cup L_2 \in \mathrm{CF}_k$.*

**Proof.** Let $X_i, Y_i \in \mathrm{CF}$ for $1 \le i \le k$ be chosen so that $L_1 = \cap_{i=1}^{k} X_i$ and $L_2 = \cap_{i=1}^{k} Y_i$. Then without loss of generality, we may assume that $X_j \subseteq R_1$ and $Y_j \subseteq R_2$ for $1 \le j \le k$; if not, we may replace $X_i$ with $X_i \cap R_1$ and $Y_i$ with $Y_i \cap R_2$ as necessary. Both intersections are still CFLs.

Thus, note that $L_1 \cup L_2 = (X_1 \cup Y_1) \cap \cdots \cap (X_k \cup Y_k)$. As $X_i \cup Y_i \in \mathrm{CF}$, the result immediately follows. ■

The following result is due to Liu and Weiner [133, Thm. 8]:

**Theorem 8.4.6** *Let $k \ge 2$. Let $L''_k = \{\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_k^{i_k}\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_k^{i_k} \ : \ i_j \ge 0\}$. Then $L''_k \in \mathrm{CF}_k - \mathrm{CF}_{k-1}$.*

However, we prove the following corollary, which will be more useful to us:

**Corollary 8.4.7** *Let $L'_k = \{\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_k^{i_k}\alpha_1^{i_1}\alpha_2^{i_2}\cdots\alpha_k^{i_k} \ : \ i_j \ge 1\}$. Then $L'_k \in \mathrm{CF}_k - \mathrm{CF}_{k-1}$.*

**Proof.** The sufficiency of $k$ intersections is obvious, by Lemma 8.4.4. We prove only the necessity of $k$ intersections. The proof is by induction. For $k = 2$, the result can be established by the pumping lemma. Let $k > 2$ and $S \subset [k]$. Denote by $L_k^{(S)}$ the language

$$L_k^{(S)} = \{\prod_{j\in S}\alpha_j^{i_j}\prod_{j\in S}\alpha_j^{i_j} \ : \ i_j \ge 1\}.$$

Further, note that

$$L_k^{(S)} \subseteq (\prod_{j\in S}\alpha_j^{+})^2.$$

Let $R_S = (\prod_{j\in S}\alpha_j^{+})^2$. Then note that $R_S \cap R_{S'} = \emptyset$ for all $S, S' \subseteq [k]$ (including the possibility that $S = [k]$) with $S \ne S'$. If $S \subset [k]$, where the inclusion is proper, then $L_k^{(S)} \in \mathrm{CF}_{k-1}$.

Assume that $L'_k$ can be expressed as the intersection of $k-1$ CFLs. We then note that

$$L''_k = L'_k \cup \bigcup_{S \subsetneq [k]} L_k^{(S)}.$$

By Lemma 8.4.5, $L''_k \in \mathrm{CF}_{k-1}$, a contradiction. This completes the proof. ■

Thus, we may state our main result:

**Theorem 8.4.8** *For all $k \geq 2$, there exists an $O(n^{2k-1})$-density bounded regular language $L_k$ such that $(\rightsquigarrow)^+(L_k)$ cannot be expressed as the intersection of $k-1$ context-free languages.*

**Proof.** Let $\Delta_k = \{\gamma_i, \alpha_i\}_{i=1}^k$. Let $h_k : \Delta_k^* \to \Delta_k^*$ be given by $h_k(\alpha_i) = h_k(\gamma_i) = \alpha_i$ for all $1 \leq i \leq k$. Let $D_k = (\rightsquigarrow)^+(L_k)$ and $R_k = \prod_{i=1}^k \alpha_i^+ \prod_{i=1}^k \gamma_i^+$. If $D_k \in \mathrm{CF}_{k-1}$, then $D_k \cap R_k \in \mathrm{CF}_{k-1}$ as well, by Lemma 8.4.4. We claim that this implies that $h_k(D_k \cap R_k)$ is in $\mathrm{CF}_{k-1}$.

Let $X_1, X_2, \ldots, X_{k-1} \in \mathrm{CF}$ be chosen so that

$$D_k \cap R_k = \cap_{i=1}^{k-1} X_i.$$

The inclusion $h_k(D_k \cap R_k) \subseteq \cap_{i=1}^{k-1} h_k(X_i)$ is easily verified. We now show the reverse inclusion. First, we may assume without loss of generality that $X_j \subseteq R_k$ for all $1 \leq j \leq k-1$. If not, let $X_j' = X_j \cap R_k$. By the closure properties of the CFLs, $X_j' \in \mathrm{CF}$, and we still have $D_k \cap R_k = \cap_{i=1}^{k-1} X_i'$.

Let $x \in \cap_{i=1}^{k-1} h_k(X_i)$. Let $y_i \in X_i$ be such that $h_k(y_i) = x$ for $1 \leq i \leq k-1$. By assumption, we can write

$$y_j = \prod_{i=1}^{k} \alpha_i^{\ell_i^{(j)}} \prod_{i=1}^{k} \gamma_i^{m_i^{(j)}}$$

for some $\ell_i^{(j)}, m_i^{(j)} \geq 1$ for $1 \leq i \leq k$ and $1 \leq j \leq k-1$. Thus, by definition of $h_k$,

$$h_k(y_j) = \prod_{i=1}^{k} \alpha_i^{\ell_i^{(j)}} \prod_{i=1}^{k} \alpha_i^{m_i^{(j)}},$$

for all $1 \leq j \leq k-1$. As $h_k(y_j) = x$, for all $1 \leq j \leq k-1$, we must have that $\ell_i^{(j)} = \ell_i^{(j')}$ and $m_i^{(j)} = m_i^{(j')}$ for $1 \leq i \leq k$ and all $1 \leq j, j' \leq k-1$. Thus $y_1 = \cdots = y_{k-1} \in \cap_{i=1}^{k-1} X_i$ and $x \in h_k(\cap_{i=1}^{k-1} X_i) = h_k(D_k \cap R_k)$. Therefore, $h_k(D_k \cap R_k) = \cap_{i=1}^{k-1} h_k(X_i)$. As $h_k(X_i)$ is a CFL for all $1 \leq i \leq k-1$, $h_k(D_k \cap R_k) \in \mathrm{CF}_{k-1}$. But now note that

$$h_k(D_k \cap R_k) = L_k',$$

by (8.12). This contradicts Corollary 8.4.7. Thus, $D_k$ cannot be expressed as the intersection of $k-1$ CFLs. ■

We now consider another example of representing non-regular languages by the iterated scattered deletion of a regular language. Let $\tilde{\Sigma}$ be a copy of $\Sigma$. Recall that $com(u)$ is the set of all words which can be obtained by permuting the letters of $u$, i.e.,

$$com(u) = \{v \in \Sigma^* \ : \ \forall a \in \Sigma, |u|_a = |v|_a\}.$$

**Theorem 8.4.9** *Let $L = \{u\tilde{v} \ : \ u \in \Sigma^*, v \in com(u)\}$. Then there exist regular languages $R_1, R_2$ such that $(\rightsquigarrow)^+(R_1) \cap R_2 = L$.*

**Proof.** Let $\hat{\Sigma}, \check{\Sigma}$ be two additional copies of $\Sigma$. Let $R_1 = (\bigcup_{a\in\Sigma}(a\hat{a}))^*(\bigcup_{a\in\Sigma}\tilde{a}\check{a})^* + \bigcup_{a\in\Sigma}\hat{a}\check{a}$. Let $R_2 = \Sigma^*\tilde{\Sigma}^*$.

We can establish, in the same manner as Theorem 8.4.8, that $(\rightsquigarrow)^+(R_1) \cap R_2 = L$. Again, the key step is to show that for all $x_1x_2 \in (\rightsquigarrow)^+(R_1)$, where $x_1 \in (\Sigma + \hat{\Sigma})^*$ and $x_2 \in (\tilde{\Sigma} + \check{\Sigma})^*$, the following equality holds for all $a \in \Sigma$:

$$|x_1|_a - |x_1|_{\hat{a}} = |x_2|_{\tilde{a}} - |x_2|_{\check{a}}.$$

This is easily established by induction. ■

Note that $L \in \text{CF}_{|\Sigma|}$. To see this, consider the language

$$L_a = \{w\tilde{u} \ : \ w, u \in \Sigma^*, |w|_a = |u|_a\}$$

for all $a \in \Sigma$. Then $L_a \in \text{CF}$, and $L = \cap_{a\in\Sigma}L_a$. The fact that $L \in \text{CF}_{|\Sigma|}$ is in contrast to the fact that the language $\{w\tilde{w} \ : \ w \in \Sigma^*\}$ is not in $\text{CF}_k$ for any $k \geq 1$, which was established by Wotschke [200]. Thus, there is a significant difference, in terms of descriptional complexity, between the language of (marked) squares and the language $L$ of (marked) "Abelian squares". We refer the reader to Jędrezejowicz and Szepietowski [93] for a discussion of $L$ versus $\{ww \ : \ w \in \Sigma^*\}$ as it relates to mildly context-sensitive families of languages and iterated shuffle.

We have the following open problem:

**Open Problem 8.4.10** *For all regular languages $R$, does there exist a $k \geq 1$ such that $(\rightsquigarrow)^+(R) \in \text{CF}_k$?*

### 8.4.2 Density and Iterated Deletion

From Theorem 8.4.8, we note that the $O(n^2)$-density set $T = i^*di^*di^*$ of trajectories yields an operation $(\rightsquigarrow_T)^*$ which does not preserve regularity. Indeed, if $L = (ba)^*(ce)^* + ac$, then

$$(\rightsquigarrow_T)^*(L) \cap b^*e^* = \{b^n e^n \ : \ n \geq 0\}.$$

Is is open whether there is a set $T$ of trajectories with $p_T(n) \in o(n^2)$ which does not preserve regularity.

We note by Theorem 8.4.8, there is an $O(n)$-density regular language $L$ such that $(\rightsquigarrow)^*(L)$ is not regular. Thus, we can ask the following open question:

**Open Problem 8.4.11** *Given a regular language $L$ such that $p_L(n) \in O(1)$, is $(\rightsquigarrow)^*(L)$ regular?*

We note that if $T = i^*di^*di^*di^*$ then using $L = (a_1a_2)^*(b_1b_2)^*(c_1c_2)^* + a_1b_1c_1$, we see that $(\rightsquigarrow_T)^*(L)$ is not a CFL. Again, we do not know if there exists a regular set $T \subseteq \{i, d\}^*$ with $p_T(n) \in o(n^3)$ such that the closure of the regular languages under $(\rightsquigarrow_T)^*$ contains non-CF languages. We summarize the best-known minimal densities in Table 8.4.2.

| | $p_T(n)$ | $p_L(n)$ |
|---|---|---|
| non-regular | $O(n^2)$ | $O(n)$ |
| non-CF | $O(n^3)$ | $O(n^2)$ |

Figure 8.2: Summary of minimum-density regular languages and regular sets of trajectories demonstrating non-closure properties for iterated deletion along trajectories.

## 8.5 Additional Closure Properties

We now consider some additional closure properties. We are motivated by an open problem of Ito and Silva [80] on the closure properties of the CSLs under iterated scattered deletion. We will require the following theorem, which can be found, in a slightly less general version, in, e.g., Salomaa [174, Thm. 9.9]:

**Theorem 8.5.1** *Let $\Sigma$ be an alphabet, and $a \notin \Sigma$. Let $s \in \Omega(\log(n))$ be any space constructible function. Then for all $L \in \text{RE}$ over $\Sigma$, there exists $L' \in \text{NSPACE}(s)$ such that $L' \subseteq a^* L$ and for all $x \in L$, there exists $i \geq 0$ such that $a^i x \in L'$.*

For all $T \subseteq \{i, d\}^*$, let $\text{suff}(T) = \{t \in \{i, d\}^* \ : \ \exists t' \in \{i, d\}^* \text{ such that } t't \in T\}$. Our stated closure property is given below:

**Theorem 8.5.2** *Let $s \in \Omega(\log(n))$ be a space-constructible function. Let $d^* i^* \subseteq T$. If there exists $L$ such that $(\rightsquigarrow_{\mathit{suff}(T)})^+(L) \in \text{RE} - \text{NSPACE}(s)$, then $\text{NSPACE}(s)$ is not closed under $(\rightsquigarrow_T)^+$.*

**Proof.** Let $L \subseteq \Sigma^*$ be a language such that $(\rightsquigarrow_{\text{suff}(T)})^+(L) \in \text{RE} - \text{NSPACE}(s)$. Let $a \notin \Sigma$ and $L_0 \subseteq a^* \left((\rightsquigarrow_{\text{suff}(T)})^+(L)\right)$ be the language in $\text{NSPACE}(s)$ described by Theorem 8.5.1. Let $L_1 = L_0 + a^*$. Clearly, $L_1 \in \text{NSPACE}(s)$. We claim that $(\rightsquigarrow_T)^+(L_1) \cap \Sigma^+ = (\rightsquigarrow_{\text{suff}(T)})^+(L)$.

($\supseteq$): Let $x \in (\rightsquigarrow_{\text{suff}(T)})^+(L)$. Then there exists $j \geq 0$ such that $a^j x \in L_1$. As $a^j \in L_1$ and $T \supseteq d^* i^* \ni d^j i^{|x|}$, $x \in a^j x \rightsquigarrow_T a^j \subseteq (\rightsquigarrow_T)^+(L_1)$.

($\subseteq$): To prove this inclusion, we prove the stronger claim that

$$(\rightsquigarrow_T)^+(L_1) \subseteq a^* (\rightsquigarrow_{\text{suff}(T)})^*(L).$$

Let $x \in (\rightsquigarrow_T)^+(L_1)$. Then there exists $j \geq 1$ such that $x \in (\rightsquigarrow_T)^j(L_1)$. The proof of our claim is by induction on $j$.

For $j = 1$, $x \in L_1 \subseteq a^*(\rightsquigarrow_{\text{suff}(T)})^+(L) + a^*$. Thus, the result clearly holds. Let $j > 1$ and assume the result holds for all natural numbers less than $j$. As $x \in (\rightsquigarrow_T)^j(L_1)$, then either $x \in (\rightsquigarrow_T)^{j-1}(L_1)$, whereby the result clearly holds by induction, or there exist $x_1, x_2 \in (\rightsquigarrow_T)^{j-1}(L_1)$ and $t \in T$ such that $x \in x_1 \rightsquigarrow_t x_2$. By induction, $x_i = a^{k_i} u_i$ for $k_i \geq 0$, and $u_i \in (\rightsquigarrow_{\text{suff}(T)})^*(L)$, for $i = 1, 2$.

There are three cases:

(a) $u_1, u_2 \in (\rightsquigarrow_{\text{suff}(T)})^+(L)$. Then $x \in x_1 \rightsquigarrow_t x_2$ implies that $t = t_1 t_2$ where $t_1$ satisfies $|t_1| = k_1$ and $|t_1|_d = k_2$. Thus, $x \in a^{k_1 - k_2}(u_1 \rightsquigarrow_{t_2} u_2)$. Let $u \in u_1 \rightsquigarrow_{t_2} u_2$ be such that $x = a^{k_1 - k_2} u$. Then note that $t_2 \in \text{suff}(T)$ and thus $u \in u_1 \rightsquigarrow_{t_2} u_2 \subseteq (\rightsquigarrow_{\text{suff}(T)})^+(L)$. Thus, $x \in a^*(\rightsquigarrow_{\text{suff}(T)})^*(L)$.

(b) $u_2 = \epsilon$. Then $x_2 = a^{k_2}$ and $x_1 = a^{k_1} u_1$ where $u_1 \in (\rightsquigarrow_{\text{suff}(T)})^*(L_1)$. Then note that necessarily, $x = a^{k_2-k_1} u_1$. Thus, $x \in a^*(\rightsquigarrow_{\text{suff}(T)})^*(L_1)$.

(c) $u_1 = \epsilon$ and $u_2 \in (\rightsquigarrow_T)^+(L)$ such that $u_2 \neq \epsilon$. Then $x_1 \rightsquigarrow_t x_2 = \emptyset$.

Thus, we have established that

$$(\rightsquigarrow_T)^+(L_1) \cap \Sigma^+ = (\rightsquigarrow_{\text{suff}(T)})^+(L). \tag{8.14}$$

Assume, contrary to what we want to prove, that $\text{NSPACE}(s)$ is closed under $(\rightsquigarrow_T)^+$. As $L_1 \in \text{NSPACE}(s)$, $(\rightsquigarrow_{\text{suff}(T)})^+(L) \in \text{NSPACE}(s)$ by (8.14). This contradicts our choice of $L$. Thus, we have established the result. ■

For scattered deletion, $T = (i+d)^*$, and thus $T = \text{suff}(T)$. Thus, if $\text{CS} = \text{NSPACE}(n)$ is closed under $(\rightsquigarrow)^+$, then for all $L \in \text{RE} - \text{CS}$, $(\rightsquigarrow)^+(L) \in \text{CS}$. The closure of CS under $(\rightsquigarrow_T)^+$ is an open problem posed by Ito and Silva [80].

## 8.6 $T$-Closure of a Language

We will now investigate the natural problem of, given $L$, finding the smallest language which is closed under $⧢_T$ and contains $L$. Classically, it is known that the smallest language containing $L$ which is closed under concatenation is $L^+$. This question has also been examined by Ito *et al.* [78, 79] and Kari and Thierrin [115] for other operations modeled by shuffle on trajectories. We will require some notions about quotients and residuals, which we discuss first.

### 8.6.1 Shuffle-$T$ Quotient

Let $T \subseteq \{0, 1\}^*$. In this section, we describe the shuffle-$T$ quotient of a language $L$ with respect to a language $L_1$, and show that if $L, L_1$ and $T$ are regular, the shuffle-$T$ quotient of $L_1$ with respect to $L$ is again regular.

Let $\Sigma$ be an alphabet and $L \subseteq \Sigma^*$. Then the *shuffle-$T$ quotient of $L$ with respect to $L_1$*, denoted

$sq_T(L; L_1)$, is given by

$$sq_T(L; L_1) = \{x \in \Sigma^* : \forall y \in L_1, y \sqcup\!\sqcup_T x \subseteq L\}.$$

For arbitrary shuffle $T = (0+1)^*$, the shuffle quotient has been examined by Câmpeanu *et al.* [21]. Ito *et al.* have examined (arbitrary) shuffle residual [79], which is given by $sq_T(L; L)$ for $T = (0+1)^*$ and insertion residual [78], which is given by $sq_T(L; L)$ for $T = 0^*1^*0^*$. Kari and Thierrin [115] have studied $k$-insertion residuals, given by $sq_T(L; L)$ for $T = 0^*1^*0^{\leq k}$ for arbitrary $k \geq 0$. Hsiao *et al.* [69] consider right residuals for more general word operations. Our main result of this section is the following:

**Theorem 8.6.1** *For all $L \subseteq \Sigma^*$, $sq_T(L; L_1) = \overline{(\overline{L} \rightsquigarrow_{\pi(T)} L_1)}$. Thus, if $L, L_1, T$ are regular, so is $sq_T(L; L_1)$, and it can be effectively constructed.*

**Proof.** Let $x \in sq_T(L; L_1)$. Assume, contrary to what we want to prove, that $x \in \overline{L} \rightsquigarrow_{\pi(T)} L_1$. Thus, there exist $t \in T$, $y \in \overline{L}$ and $z \in L_1$ such that $x \in y \rightsquigarrow_{\pi(t)} z$. By Theorem 5.8.2, $y \in z \sqcup\!\sqcup_t x$. As $x \in sq_T(L; L_1)$ and $z \in L_1$, $z \sqcup\!\sqcup_t x \subseteq L$. However, $y \in \overline{L}$, a contradiction.

Let $x \notin sq_T(L; L_1)$. Thus, there exists some $u \in L_1$ such that $u \sqcup\!\sqcup_T x \cap \overline{L} \neq \emptyset$. Let $y$ be some word in this intersection, and let $t \in T$ be such that $y \in u \sqcup\!\sqcup_t x$. Thus by Theorem 5.8.2, $x \in y \rightsquigarrow_{\pi(t)} u \subseteq \overline{L} \rightsquigarrow_{\pi(T)} L_1$. Thus, we conclude that $\overline{sq_T(L; L_1)} \subseteq \overline{L} \rightsquigarrow_{\pi(T)} L_1$.

The fact that the regular languages are effectively closed under deletion along regular trajectories implies that $sq_T(L; L_1)$ is a regular language. This completes the proof. ■

Note that Theorem 8.6.1 gives an alternate proof that if $L_1, L_2$ are regular, then the (arbitrary) shuffle quotient of $L_1$ and $L_2$ is regular (this was originally proven by Câmpeanu *et al.* [21, Lemma 4]). Further, Theorem 8.6.1 was proven for $L = L_1$ and $T = (0+1)^*$ by Ito *et al.* [79, Prop. 2.4], for $L = L_1$ and $T = 0^*1^*0^*$ by Ito *et al.* [78, Prop. 2.3], and for $L = L_1$ and $T = 0^*1^*0^{\leq k}$ for fixed $k \geq 0$ by Kari and Thierrin [115, Prop. 2.3]. An equivalent formulation of Theorem 8.6.1 was given by Hsiao *et al.* [69, Prop. 30], but without explicitly using the notion of inverse operations. Further, in their framework of word operations, we cannot conclude any closure properties.

### 8.6.2 $T$-closure

Let $r_T(L) = sq_T(L; L)$, which we call the *shuffle-$T$ residual* of $L$. A language $L \subseteq \Sigma^*$ such that $L \subseteq r_T(L)$ is said to be *shuffle-$T$ closed*. Define

$$C_T(L) = \{L' \subseteq \Sigma^* \ : \ L \subseteq L' \subseteq r_T(L')\}.$$

Then $C_T(L)$ is the set of all shuffle-$T$ closed languages containing $L$ ($C_T(L) \neq \emptyset$ as $\Sigma^* \in C_T(L)$). Further, define

$$cl_T(L) = \bigcap_{L' \in C_T(L)} L'.$$

Then $cl_T(L)$ is the smallest $T$-closed language containing $L$; we call $cl_T(L)$ the *shuffle-$T$ closure of $L$*.

**Proposition 8.6.2** *Let $T \subseteq \{0,1\}^*$. Then $L$ is shuffle-$T$ closed if and only if $L \sqcup\!\sqcup_T L \subseteq L$.*

**Proof.** Let $L$ be shuffle-$T$ closed. Then for all $x \in L$, we have $x \in r_T(L)$. Thus, for all $u \in L$, $u \sqcup\!\sqcup_T x \subseteq L$. Clearly then, $L \sqcup\!\sqcup_T L \subseteq L$.

For the reverse implication, let $L \sqcup\!\sqcup_T L \subseteq L$. Then let $x \in L$; we show $x \in r_T(L)$. As $L \sqcup\!\sqcup_T L \subseteq L$, for all $y \in L$, $y \sqcup\!\sqcup_T x \subseteq L$. Thus, by definition, $x \in r_T(L)$. ■

Proposition 8.6.2 was noted for scattered deletion by Ito *et al.* [79], for sequential deletion by Ito *et al.* [78] and for $k$-deletion by Kari and Thierrin [115]. For catenation, a weakened version of the only-if portion of the result is sometimes given as an easy undergraduate exercise (see, e.g., Martin [141, Ex. 2.22]). In the framework of general word operations, a variant of Proposition 8.6.2 is given by Hsiao *et al.* [69, Prop. 24].

**Corollary 8.6.3** *Let $T \subseteq \{0,1\}^*$ be a regular set of trajectories. Given a regular language $L$, it is decidable whether $L$ is shuffle-$T$ closed.*

We now seek to give a characterization of $cl_T(L)$ for all $T \subseteq \{0,1\}^*$. The following fact is obvious from the definition of $(\sqcup\!\sqcup_T)^i$:

**Fact 8.6.4** *For all* $L \subseteq \Sigma^*$ *and all* $i \geq 1$, $(\sqcup\!\sqcup_T)^i(L) \subseteq (\sqcup\!\sqcup_T)^{i+1}(L)$.

**Theorem 8.6.5** *Let* $T \subseteq \{0,1\}^*$. *Then* $cl_T(L) = (\sqcup\!\sqcup_T)^+(L)$.

**Proof.** Note that $L = (\sqcup\!\sqcup_T)^1(L) \subseteq (\sqcup\!\sqcup_T)^+(L)$. To show that $cl_T(L) \subseteq (\sqcup\!\sqcup_T)^+(L)$, it suffices to show that $(\sqcup\!\sqcup_T)^+(L)$ is shuffle-$T$ closed. We appeal to Proposition 8.6.2. In particular, we show that $(\sqcup\!\sqcup_T)^+(L) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^+(L) \subseteq (\sqcup\!\sqcup_T)^+(L)$. Let $x, y \in (\sqcup\!\sqcup_T)^+(L)$. Then there exist $j, k \geq 1$ such that $x \in (\sqcup\!\sqcup_T)^j(L)$ and $y \in (\sqcup\!\sqcup_T)^k(L)$. Let $m = \max(j,k)$. Then clearly $x, y \in (\sqcup\!\sqcup_T)^m(L)$. Thus, by definition of $(\sqcup\!\sqcup_T)^m(L)$,

$$x \sqcup\!\sqcup_T y \subseteq (\sqcup\!\sqcup_T)^{m+1}(L) \subseteq (\sqcup\!\sqcup_T)^+(L).$$

The inclusion is proven.

We now show that $(\sqcup\!\sqcup_T)^+(L) \subseteq cl_T(L)$. Again, the proof is by induction on $i$: we show $(\sqcup\!\sqcup_T)^i(L) \subseteq cl_T(L)$ for all $i \geq 1$.

For $i = 1$, $L \subseteq cl_T(L)$ by definition of $cl_T(L)$. Now let $i > 1$ and assume the result holds for all integers less than $i$. Consider

$$\begin{aligned}
(\sqcup\!\sqcup_T)^i(L) &= \left((\sqcup\!\sqcup_T)^{i-1}(L) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^{i-1}(L)\right) + (\sqcup\!\sqcup_T)^{i-1}(L) \\
&\subseteq (cl_T(L) \sqcup\!\sqcup_T cl_T(L)) + cl_T(L) \\
&\subseteq cl_T(L) + cl_T(L) = cl_T(L)
\end{aligned}$$

where the first inclusion is by induction on $i$, and the second inclusion is by the fact that $cl_T(L)$ is $T$-closed (by definition of $cl_T(L)$), and Proposition 8.6.2. ■

Theorem 8.6.5 was also proven for sequential insertion by Ito *et al.* [78, Prop. 2.4], and for arbitrary shuffle by Ito *et al.* [79, Prop. 2.6].

Recall that $(\sqcup\!\sqcup_T)^*_X$ is the iterated version of $\sqcup\!\sqcup_T$ defined by (8.4). As we have stated, it is not hard to establish that $(\sqcup\!\sqcup_T)^*_X(L) = (\sqcup\!\sqcup_T)^*(L)$ for all $L$ if $T$ is left-associative. Thus, we can conclude that if $T$ is left-associative, $cl_T(L) = (\sqcup\!\sqcup_T)^+_X(L)$ for all $L$. We now show that the requirement that $T$ be left-associative is necessary for $cl_T(L) = (\sqcup\!\sqcup_T)^+_X(L)$.

**Lemma 8.6.6** *There exist a singleton language $L$ and set $T$ of non-left-associative trajectories such that $cl_T(L) \neq (\sqcup\!\sqcup_T)^+_X(L)$.*

**Proof.** We show, in fact, that infinitely many pairs $(L, T)$ exist satisfying the lemma. Let $k \geq 1$, and $T_k = 0^*1^*0^{\leq k}$. Then $\sqcup\!\sqcup_{T_k} = \overset{k}{\leftarrow}$, the $k$-insertion operation, studied by Kari and Thierrin [115]. For each $k \geq 1$, we define $L_k = \{ba^k\}$. Then we claim that

$$cl_{T_k}(L_k) \neq (\overset{k}{\leftarrow})^+_X(L_k). \tag{8.15}$$

We establish first that

$$\{b^i a^{ki} : i \geq 1\} \subseteq cl_{T_k}(L_k).$$

As $L_k \subseteq cl_{T_k}(L_k)$, $ba^k \in cl_{T_k}(L_k)$. For each $i > 1$, $ba^k, b^i a^{ki} \in cl_{T_k}(L_k)$ imply that $b^{i+1}a^{k(i+1)} \in ba^k \overset{k}{\leftarrow} b^i a^{ki} \subseteq cl_{T_k}(L_k)$.

Now, we note that $b^3(a+b)^* \cap (\overset{k}{\leftarrow})^+_X(L_k) = \emptyset$. To see this, note that if $x \in (\overset{k}{\leftarrow})^i_X(L_k)$, then $|x| = ki$ and $|x|_b = i$. We can then prove, by induction, that at most 2 occurrences of $b$ can occur at the start of any word in $(\overset{k}{\leftarrow})^i_X(L_k)$, since $k$-insertions always happen "close to the right end" of the word. Thus, we can establish (8.15). ■

Note that Lemma 8.6.6 corrects an error in Kari and Thierrin [115, Prop. 2.4], where it is claimed that $cl_{T_k}(L) = (\sqcup\!\sqcup_{T_k})^+_X(L)$ for each $T_k = 0^*1^*0^{\leq k}$.

### 8.6.3 Codes and Shuffle-Closed Languages

We now show that $T$-codes are shuffle-$T$ closed if and only if they are trivially shuffle-$T$ closed.

**Lemma 8.6.7** *Let $T \subseteq \{0,1\}^*$. Let $L \in \mathcal{P}_T(\Sigma)$. Then $L$ is shuffle-$T$ closed if and only if $L \sqcup\!\sqcup_T L = \emptyset$.*

**Proof.** If $L \sqcup\!\sqcup_T L = \emptyset$, then clearly $L \sqcup\!\sqcup_T L \subseteq L$, and $L$ is shuffle-$T$ closed.

Let $L \in \mathcal{P}_T(\Sigma)$. Then note that necessarily $L \subseteq \Sigma^+$. Let $L$ be shuffle-$T$ closed. Assume there exist $x, y \in L$ such that $x \sqcup\!\sqcup_T y \neq \emptyset$. Let $z \in x \sqcup\!\sqcup_T y$. Thus, $z \in L$ as $L$ is shuffle-$T$ closed. Therefore, $z \in L \cap (L \sqcup\!\sqcup_T \Sigma^+)$, contradicting that $L$ is a $T$-code. ■

**Corollary 8.6.8** *Let $T \subseteq \{0,1\}^*$ be complete. Let $L \in \mathcal{P}_T(\Sigma)$. Then $L$ is shuffle-$T$ closed if and only if $L = \emptyset$.*

## 8.7 Deletion Closure of a Language

We now consider the problem of, given a language $L$, finding the smallest language which contains $L$ and is closed under $\rightsquigarrow_T$.

### 8.7.1 Del-$T$ Quotient

Let $T \subseteq \{i,d\}^*$. In this section, we describe the del-$T$ quotient of a language with respect to a language $L$, and show that if $L, L_1, T$ are regular, the del-$T$ quotient of $L_1$ with respect to $L$ is again regular.

We define the set of *$T$-scattered subwords* as follows:

$$scs_T(L) = L \rightsquigarrow_{sym_d(T)} \Sigma^*.$$

Note that

$$scs_T(L) = \{u \in \Sigma^* \ : \ \exists v \in \Sigma^* \text{ such that } u \sqcup\!\sqcup_{\pi^{-1}(T)} v \cap L \neq \emptyset\}.$$

Further, note that if $L, T$ are regular, then $scs_T(L)$ is regular. As examples, note that

(a) when $T = i^*d^*i^*$, we have $scs_T(L) = sub(L)$, the subwords of $L$ (e.g., Ito *et al.* [78]), given by

$$sub(L) = \{u \in \Sigma^* \ : \ \exists x, y \in \Sigma^* \text{ such that } xuy \in L\}.$$

(b) if $T = (i+d)^*$, we have that $scs_T(L) = sps(L)$, the scattered (or sparse) subwords of $L$ (e.g., Ito *et al.* [79]), given by

$$sps(L) = \{u \in \Sigma^* \ : \ \exists v \in \Sigma^* \text{ such that } u \sqcup\!\sqcup v \cap L \neq \emptyset\}.$$

Let $\Sigma$ be an alphabet and $L \subseteq \Sigma^*$. Then the *deletion-$T$ quotient of $L$ with respect to $L_1$*, denoted $dq_T(L; L_1)$, is given by

$$dq_T(L; L_1) = \{x \in scs_T(L) \ : \ \forall y \in L_1, y \rightsquigarrow_T x \subseteq L\}.$$

Ito *et al.* have examined scattered deletion residual [79], which is given by $dq_T(L; L)$ for $T = (i + d)^*$ and deletion residual [78], which is given by $dq_T(L; L)$ for $T = i^*d^*i^*$. For $k \geq 1$, Kari and Thierrin [115] have studied $k$-deletion residual, which is given by $dq_T(L; L)$ for $T = i^*d^*i^{\leq k}$. Note that we could also define

$$dq'_T(L; L_1) = \{x \in \Sigma^* \ : \ \forall y \in L_1, y \rightsquigarrow_T x \subseteq L\}.$$

In this case, we get

$$dq'_T(L; L_1) = dq_T(L; L_1) \cup \overline{scs_T(L)}.$$

Our main result of the section is the following:

**Theorem 8.7.1** *For all $L \subseteq \Sigma^*$, $dq_T(L; L_1) = \overline{(L_1 \rightsquigarrow_{sym_d(T)} \overline{L})} \cap scs_T(L)$. Thus, if $L, L_1, T$ are regular, so is $dq_T(L; L_1)$, and it can be effectively constructed.*

**Proof.** Let $x \in dq_T(L; L_1)$. Immediately, we have that $x \in scs_T(L)$. Assume, contrary to what we want to prove, that $x \in L_1 \rightsquigarrow_{sym_d(T)} \overline{L}$. Thus, there exist $t \in T$, $y \in \overline{L}$ and $z \in L_1$ such that $x \in z \rightsquigarrow_{sym_d(t)} y$. By Theorem 5.8.3, $y \in z \rightsquigarrow_t x$. As $x \in dq_T(L; L_1)$ and $z \in L_1$, $z \rightsquigarrow_t x \subseteq L$. However, $y \in \overline{L}$, a contradiction.

Let $x \in \overline{L_1 \rightsquigarrow_{sym_d(T)} \overline{L}} \cap scs_T(L)$. Assume, contrary to what we want to prove, that $x \notin dq_T(L; L_1)$. As $x \in scs_T(L)$, this implies that there exists a word $y \in L_1$ such that $y \rightsquigarrow_T x \cap \overline{L} \neq \emptyset$. Let $u$ be some word in this intersection. Thus, there is some $t \in T$ such that $u \in y \rightsquigarrow_t x$. By Theorem 5.8.3, $x \in y \rightsquigarrow_{sym_d(t)} u \subseteq L_1 \rightsquigarrow_{sym_d(T)} \overline{L}$. This contradicts our choice of $x$. ■

Theorem 8.7.1 was proven by Ito *et al.* for the cases where $L = L_1$ and $T = (i + d)^*$ [79, Prop. 4.2] as well as $L = L_1$ and $T = i^*d^*i^*$ [78, Prop. 3.2]. Theorem 8.7.1 was proven by Kari and Thierrin [115] for the case of $L = L_1$ and $T = i^*d^*i^{\leq k}$ for fixed $k \geq 1$.

### 8.7.2 $T$-del-closure

Let $T \subseteq \{i, d\}^*$ Let $dr_T(L) = dq_T(L; L)$, which we call the *del-$T$ residual* of $L$. A language $L$ such that $L \cap scs_T(L) \subseteq dr_T(L)$ is said to be *del-$T$ closed*.

We first note a class of trajectories for which the annoyance of dealing with $scs_T(L)$ is removed. We call a set $T \subseteq \{i, d\}^*$ *del-left-preserving* with respect to $L$ if $i^{|x|} \in T$ for all $x \in L$. Note that if $T$ is del-left-preserving with respect to every language $L$ then $T$ is del-left-preserving. If $sym_d(T)$ is del-left-preserving (with respect to $L$), we say that $T$ is *sym-del-left-preserving* (with respect to $L$), or *sdl-preserving*.

**Lemma 8.7.2** *Let $L \subseteq \Sigma^*$. If $T$ is sdl-preserving with respect to $L$, then $L \subseteq scs_T(L)$.*

**Proof.** Note in this case that $scs_T(L) = L \rightsquigarrow_{sym_d(T)} \Sigma^* \supseteq L \rightsquigarrow_{sym_d(T)} \{\epsilon\} = L$, as $sym_d(T)$ is del-left-preserving. ■

Note that if $T$ is sdl-preserving, $T \supseteq d^*$. This is satisfied by, for example, right- and left-quotient, and sequential, bi-polar, $k$- and scattered deletion.

Consider that if $L \subseteq scs_T(L)$, then clearly $L = scs_T(L) \cap L$. This leads to the following observation:

**Proposition 8.7.3** *Let $T$ be sdl-preserving with respect to $L$. Then $L$ is del-$T$ closed if and only if $L \subseteq dr_T(L)$.*

For defining the $T$-del-closure of a language, we need the following notation. Define

$$dC_T(L) = \{L' \subseteq \Sigma^* \ : \ L \subseteq L' \text{ and } L' \cap scs_T(L') \subseteq dr_T(L')\}.$$

Then $dC_T(L)$ is the set of all del-$T$ closed languages containing $L$ ($dC_T(L) \neq \emptyset$ as $\Sigma^* \in dC_T(L)$). Further, define $dcl_T(L) = \cap_{L' \in dC_T(L)} L'$. It is not hard to see that

$$scs_T(dcl_T(L)) \subseteq \bigcap_{L' \in dC_T(L)} scs_T(L'), \tag{8.16}$$

and that

$$dcl_T(L) \cap scs_T(dcl_T(L)) \subseteq \bigcap_{L' \in dC_T(L)} dr_T(L). \tag{8.17}$$

With this, we can see that $dcl_T(L)$ is the smallest $T$-del-closed language containing $L$; we call $dcl_T(L)$ the *del-$T$ closure of $L$*.

**Proposition 8.7.4** *Let $T \subseteq \{i,d\}^*$. Then $L$ is del-$T$ closed if and only if $L \rightsquigarrow_T (L \cap scs_T(L)) \subseteq L$.*

**Proof.** The proof is similar to that of Lemma 8.6.2. Let $L$ be del-$T$ closed. Then for all $x \in L \cap scs_T(L)$, we have $x \in dr_T(L)$. Thus, for all $u \in L$, $u \rightsquigarrow_T x \subseteq L$. Clearly then, $L \rightsquigarrow_T (L \cap scs_T(L)) \subseteq L$.

For the reverse implication, let $L \rightsquigarrow_T (L \cap scs_T(L)) \subseteq L$. Consider $x \in L \cap scs_T(L)$; we show $x \in dr_T(L)$. As $L \rightsquigarrow_T (L \cap scs_T(L)) \subseteq L$, for all $y \in L$, $y \rightsquigarrow_T x \subseteq L$. Thus, by definition, as $x \in scs_T(L)$, we also have $x \in dr_T(L)$. ■

**Corollary 8.7.5** *Let $L \subseteq \Sigma^*$. Let $T \subseteq \{i,d\}^*$ be sdl-preserving with respect to $L$. Then $L$ is del-$T$-closed if and only if $L \rightsquigarrow L \subseteq L$.*

Corollary 8.7.5 was noted by Ito *et al.* for $T = (i+d)^*$ [79] and $T = i^*d^*i^*$ [78]. We can also generalize an interesting result of Ito *et al.* [78, 79] and Kari and Thierrin [115, Prop. 3.3]. Call a set of trajectories $T$ *square-enabling* if $\Psi(T) \supseteq \{(n,n) : n \geq 0\}$.

**Lemma 8.7.6** *Let $T \subseteq \{0,1\}^*$ be square-enabling, and such that $\tau(T)$ is sdl-preserving. Let $L$ be a shuffle-$T$ closed language. Then $L$ is del-$\tau(T)$ closed if and only if $L = L \rightsquigarrow_{\tau(T)} L$.*

**Proof.** If $L = L \rightsquigarrow_{\tau(T)} L$, then by Corollary 8.7.5, $L$ is $\tau(T)$-del-closed.

Now, assume that $L$ is $\tau(T)$-del-closed. Again by Corollary 8.7.5, $L \supseteq L \rightsquigarrow_{\tau(T)} L$. Thus, let $x \in L$. we must show $x \in L \rightsquigarrow_{\tau(T)} L$.

As $L$ is shuffle-$T$ closed, $x \sqcup\!\sqcup_T x \subseteq L$. As $T$ is square-enabling, $x \sqcup\!\sqcup_T x \neq \emptyset$. Thus, let $t \in T$ and $y \in L$ be chosen so that $y \in x \sqcup\!\sqcup_t x$. By Theorem 5.8.2, $x \in y \rightsquigarrow_{\tau(t)} x$. Thus, $x \in L \rightsquigarrow_{\tau(T)} L$, as required. ■

Let $T \subseteq \{i, d\}^*$. Let $L$ be a language. We now give a characterization of the $T$-del-closure of a language $L$, when $T$ is sdl-preserving.

**Fact 8.7.7** *For all $T \subseteq \{i, d\}^*$ and $L \subseteq \Sigma^*$, $(\rightsquigarrow_T)^i(L) \subseteq (\rightsquigarrow_T)^{i+1}(L)$.*

**Theorem 8.7.8** *Let $L \subseteq \Sigma^*$ be a language and $T \subseteq \{i, d\}^*$ be sdl-preserving. Then $dcl_T(L) = (\rightsquigarrow_T)^+(L)$.*

**Proof.** Note that $L \subseteq (\rightsquigarrow_T)^+(L)$. Then to show that $dcl_T(L) \subseteq (\rightsquigarrow_T)^+(L)$, it suffices to show that $(\rightsquigarrow_T)^+(L)$ is del-$T$ closed. We appeal to Corollary 8.7.5. In particular, we show that

$$(\rightsquigarrow_T)^+(L) \rightsquigarrow_T (\rightsquigarrow_T)^+(L) \subseteq (\rightsquigarrow_T)^+(L).$$

Let $u, v \in (\rightsquigarrow_T)^+(L)$. Then there exist $i, j \geq 1$ such that $u \in (\rightsquigarrow_T)^i(L)$ and $v \in (\rightsquigarrow_T)^j(L)$. Let $k = \max(i, j)$. This implies that $u, v \in (\rightsquigarrow_T)^k(L)$. Thus, $u \rightsquigarrow v \subseteq (\rightsquigarrow_T)^{k+1}(L)$ by definition of $(\rightsquigarrow_T)^k$. We conclude that $(\rightsquigarrow_T)^+$ is del-$T$ closed and thus $dcl_T(L) \subseteq (\rightsquigarrow_T)^+(L)$.

To show the reverse inclusion, we show by induction on $i$ that $(\rightsquigarrow_T)^i(L) \subseteq dcl_T(L)$ for all $i \geq 1$. For $i = 1$, $L \subseteq dcl_T(L)$ by definition of $dcl_T(L)$. Now assume the result holds for all natural numbers less than $i$. Consider

$$\begin{aligned} (\rightsquigarrow_T)^i(L) &= \left((\rightsquigarrow_T)^{i-1}(L) \rightsquigarrow_T (\rightsquigarrow_T)^{i-1}(L)\right) + (\rightsquigarrow_T)^{i-1}(L); \\ &\subseteq (dcl_T(L) \rightsquigarrow_T dcl_T(L)) + dcl_T(L); \\ &\subseteq (dcl_T(L)) + dcl_T(L) = dcl_T(L), \end{aligned}$$

where the first inclusion is by induction on $i$, and the second inclusion is by Corollary 8.7.5, and the fact that $dcl_T(L)$ is $T$-del-closed. ■

Theorem 8.7.8 was proven by Ito *et al.* in the case of scattered [79, Prop. 4.4] and sequential [78, Prop. 3.4] deletion. Theorem 8.7.8 was proven by Kari and Thierrin [115, Prop. 3.4] for the case of $k$-deletion.

We now show that sdl-preservation is necessary in the alternate definition of $(\rightsquigarrow_T)^+_X(L)$ given by (8.9):

**Lemma 8.7.9** *There exist a set of trajectories $T$ and infinitely many languages $L$ such that $T$ is not sdl-preserving with respect to $L$ and $dcl_T(L) \neq (\rightsquigarrow_T)_X^+(L)$.*

**Proof.** We introduce a rather extreme example of a set $T$ satisfying the conditions. Let $T = d^*$. Then note that unless $L \subseteq \{\epsilon\}$, $T$ is not sdl-preserving with respect to $L$. Further, note that

$$L_1 \rightsquigarrow_T L_2 = \begin{cases} \{\epsilon\} & \text{if } L_1 \cap L_2 \neq \emptyset. \\ \emptyset & \text{otherwise.} \end{cases}$$

Then let $L$ be any language such that $\{\epsilon\}$ is properly contained in $L$. Then $L \rightsquigarrow_T L = \{\epsilon\}$ and by induction, we can show that $(\rightsquigarrow_T)_X^i(L) \rightsquigarrow_T ((\rightsquigarrow_T)_X^i(L) + \epsilon) = \{\epsilon\}$, for all $i \geq 2$. Thus, $(\rightsquigarrow_T)_X^+(L) = \{\epsilon\}$. But clearly in this case, $dcl_T(L) \neq \{\epsilon\}$, since $L \subseteq dcl_T(L)$ by definition. ■

## 8.8 $T$-Shuffle Base

We now extend the notion of shuffle base, examined by Ito *et al.* [78, 79] and Kari and Thierrin [115]. Let $L \subseteq \Sigma^+$ be a shuffle-$T$-closed language (the following definitions can be given for $L \subseteq \Sigma^*$, as was done by Ito *et al.* [78, 79] and Kari and Thierrin [115], but we restrict this possibility to allow for simpler definitions; the results below can also be given for the more complete definitions of Ito *et al.* and Kari and Thierrin without much difficulty). The *shuffle-$T$ base* of $L$, denoted by $J_T(L)$, is the set of words in $L$ which cannot be expressed as the shuffle along $T$ of words in $L$. Thus,

$$J_T(L) = \{u \in L \ : \ u \notin L \sqcup\!\sqcup_T L\}.$$

**Proposition 8.8.1** *Let $T$ be left-associative and $L \subseteq \Sigma^+$ be shuffle-$T$ closed. Then*

$$L = (\sqcup\!\sqcup_T)^+(J_T(L)). \tag{8.18}$$

**Proof.** As $L$ is shuffle-$T$ closed, $L = (\sqcup\!\sqcup_T)^+(L)$. Thus, it suffices to show that $(\sqcup\!\sqcup_T)^+(L) = (\sqcup\!\sqcup_T)^+(J_T(L))$. As $T$ is left-associative, we reduce this to showing following equality:

$$(\sqcup\!\sqcup_T)_X^+(L) = (\sqcup\!\sqcup_T)_X^+(J_T(L)).$$

To see that $(\sqcup\!\sqcup_T)^+_X(J_T(L)) \subseteq (\sqcup\!\sqcup_T)^+_X(L)$, we note that $J_T(L) \subseteq L$ and $(\sqcup\!\sqcup_T)^+_X$ is a monotone operator, since $\sqcup\!\sqcup_T$ is. For the reverse inclusion, let $x \in (\sqcup\!\sqcup_T)^+_X(L)$. Then we note that as $\epsilon \notin L$, if $x \in (\sqcup\!\sqcup_T)^i_X(L)$, then $i \leq |x|$. Thus, let $h_x$ be the maximum $i$ such that $x \in (\sqcup\!\sqcup_T)^i_X(L)$. We prove that $x \in (\sqcup\!\sqcup_T)^+_X(J_T(L))$ by induction on $h_x$.

For $h_x = 1$, $x \in L$ and $x \notin (\sqcup\!\sqcup_T)^2_X(L) = L \sqcup\!\sqcup_T L$. Thus, by definition, $x \in J_T(L) \subseteq (\sqcup\!\sqcup_T)^+_X(J_T(L))$. Assume the result holds for all $x$ with $h_x < h$ for some $h > 1$. Let $x$ be a word such that $h_x = h$. Then $x \in (\sqcup\!\sqcup_T)^h_X(L)$, and there exist $y \in (\sqcup\!\sqcup_T)^{h-1}_X(L)$ and $z \in L$ such that $x \in y \sqcup\!\sqcup_T z$. By induction, $y \in (\sqcup\!\sqcup_T)^+_X(J_T(L))$. If $z \in J_T(L)$, we are done. Therefore, assume that $z \in L - J_T(L)$. There exist $u, v \in L$ such that $z = u \sqcup\!\sqcup_T v$. Thus, $x \in y \sqcup\!\sqcup_T (u \sqcup\!\sqcup_T v) \subseteq (y \sqcup\!\sqcup_T u) \sqcup\!\sqcup_t v$, as $T$ is left-associative. But now $y \in (\sqcup\!\sqcup_T)^{h-1}_X(L)$, $u, v \in L$ imply that $x \in (\sqcup\!\sqcup_T)^{h+1}_X(L)$, a contradiction to our choice of $x$. Thus, $z \in J_T(L)$ and $x \in (\sqcup\!\sqcup_T)^+_X(J_T(L))$. ■

We now demonstrate that if $L$ and $T$ are regular and $L$ is shuffle-$T$ closed, then $J_T(L)$ is also regular.

**Lemma 8.8.2** *Let $T \subseteq \{0,1\}^*$ be a regular set of trajectories. If $L \subseteq \Sigma^+$ is regular and shuffle-$T$-closed, then $J_T(L)$ is regular.*

**Proof.** The proof will rely on establishing that

$$L - J_T(L) = L \sqcup\!\sqcup_T L. \tag{8.19}$$

Let $x \in L - J_T(L)$. Then as $x \notin J_T(L)$, there exist $u, v \in L$ such that $x \in u \sqcup\!\sqcup_T v$. Thus $x \in L \sqcup\!\sqcup_T L$. Now, let $x \in L \sqcup\!\sqcup_T L$. As $L$ is shuffle-$T$-closed, $L \sqcup\!\sqcup_T L \subseteq L$. Thus $x \in L$. As $x \in L \sqcup\!\sqcup_T L$, $x \notin J_T(L)$ by definition. This establishes (8.19). Now, as $L \sqcup\!\sqcup_T L, J_T(L) \subseteq L$, we have that $J_T(L) = L - L \sqcup\!\sqcup_T L$. Since $L$ and $L \sqcup\!\sqcup_T L$ are regular, $J_T(L)$ is regular. ■

Lemma 8.8.2 offers alternate proofs of the corresponding results by Ito *et al.* for scattered deletion [79, Prop. 5.1] and sequential deletion [78, Prop. 5.1], and by Kari and Thierrin for $k$-insertion [115, Prop. 2.5].

## 8.9 Shuffle-Free Languages

In this section, we consider the notion of shuffle-free languages. This was first examined by Ito *et al.* [82]. Hsiao *et al.* [69, Sect. 5] also considered a similar notion for arbitrary word operations. We show that for shuffle on trajectories, we can obtain results relating shuffle-free languages and $T$-codes.

We adopt the following shorthand:

$$(\sqcup\!\sqcup_T)^{\geq 2}(L) = \bigcup_{i \geq 2} (\sqcup\!\sqcup_T)^i(L).$$

Let $T \subseteq \{0,1\}^*$. Then we say that $\emptyset \neq L \subseteq \Sigma^+$ is *sh-$T$-free* if

$$(\sqcup\!\sqcup_T)^{\geq 2}(L) \cap L = \emptyset.$$

Thus, $L$ is sh-$T$-free if, for all $u_1, u_2, \ldots, u_k \in L$ ($k \geq 2$), there is no way to shuffle the $u_j$s together with $\sqcup\!\sqcup_T$ to get a word from $L$. The concept of sh-$T$-free is an extension of the corresponding notion, sh-free, for arbitrary shuffle; this was introduced by Ito *et al.* [82].

The following lemma will be useful:

**Lemma 8.9.1** *Let $T \subseteq \{0,1\}^*$. Let $L_1, L_2$ be two sh-$T$-free languages such that $(\sqcup\!\sqcup_T)^+(L_1) = (\sqcup\!\sqcup_T)^+(L_2)$. Then $L_1 = L_2$.*

**Proof.** Assume not. Let $x \in L_1 - L_2$, without loss of generality. Since $x \in L_1 \subseteq (\sqcup\!\sqcup_T)^+(L_1) = (\sqcup\!\sqcup_T)^+(L_2)$, either $x \in L_2$ or there exist $u_1, u_2 \in (\sqcup\!\sqcup_T)^+(L_2)$ such that $x \in u_1 \sqcup\!\sqcup_T u_2$. As $x \notin L_2$ by assumption, let $x \in u_1 \sqcup\!\sqcup_T u_2$. We now note that $u_1, u_2 \in (\sqcup\!\sqcup_T)^+(L_2) = (\sqcup\!\sqcup_T)^+(L_1)$. Thus, $x \in (\sqcup\!\sqcup_T)^+(L_1) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^+(L_1) \subseteq (\sqcup\!\sqcup_T)^{\geq 2}(L_1)$. This contradicts that $L_1$ is sh-$T$-free. ■

Let $L \subseteq \Sigma^*$. We say that $B_T(L) \subseteq L \cap \Sigma^+$ is an *extended sh-$T$-base* of $L$ if $B_T(L)$ is sh-$T$-free and $L \cap \Sigma^+ \subseteq (\sqcup\!\sqcup_T)^+(B_T(L))$.

**Lemma 8.9.2** *Let $T \subseteq \{0,1\}^*$ and let $L \subseteq \Sigma^*$. Then the extended sh-$T$-base of $L$ is unique.*

**Proof.** Let $B_1, B_2$ be two extended sh-$T$-bases of $L$.

Let $x \in B_1$. Then $x \neq \epsilon$ by definition. Further, $x \in \Sigma^+ \cap L \subseteq (\sqcup\!\sqcup_T)^+(B_2)$, by the fact that $B_2$ is an extended-$T$-base of $L$. Thus, $B_1 \subseteq (\sqcup\!\sqcup_T)^+(B_2)$, and

$$(\sqcup\!\sqcup_T)^+(B_1) \subseteq (\sqcup\!\sqcup_T)^+((\sqcup\!\sqcup_T)^+(B_2)) \subseteq (\sqcup\!\sqcup_T)^+(B_2),$$

where the last inclusion is valid by Theorem 8.6.5. By symmetry, we also have that $(\sqcup\!\sqcup_T)^+(B_2) \subseteq (\sqcup\!\sqcup_T)^+(B_1)$.

As $B_1, B_2$ are sh-$T$-free, we have that $B_1 = B_2$, by Lemma 8.9.1. ■

We now show that every language has a sh-$T$-base. Consider the following languages [69, 82]:

$$\begin{aligned} B_0 &= \emptyset; \\ K_i &= L - (\sqcup\!\sqcup_T)^+(\bigcup_{j=0}^{i-1} H_j); \quad \forall i \geq 1 \\ B_i &= \{x \ : \ x \in K_i \text{ and } |x| \leq |y| \quad \forall y \in K_i\}; \quad \forall i \geq 1 \\ B &= \bigcup_{i \geq 1} B_i. \end{aligned}$$

Then it is straight-forward to establish that $B$ is a sh-$T$-base for $L$ (see, e.g., Hsiao *et al.* [69, Prop. 12]).

There is an interesting relation between extended sh-$T$-bases and $T$-codes. This emphasizes the naturalness of the definition of $T$-codes.

**Lemma 8.9.3** *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that $\pi(T)$ is sdl-preserving. If $B_T(L)$ is the extended sh-$T$-base of any del-$\pi(T)$ closed language $L$, then $B_T(L)$ is a $T$-code.*

**Proof.** Let $B_T(L)$ be the extended sh-$T$-base of $L$. Suppose $B_T(L)$ is not a $T$-code. Then there exist $u, v \in B_T(L) \subseteq L$ such that $v \in u \sqcup\!\sqcup_T w$ for some $w \in \Sigma^+$. By Theorem 5.8.2, $w \in v \rightsquigarrow_{\pi(T)} u$.

As $u, v \in L$, $w \in L$ by Corollary 8.7.5. Thus, $w \in L \cap \Sigma^+ \subseteq (\sqcup\!\sqcup_T)^+(B_T(L))$. But then $v \in u \sqcup\!\sqcup_T w \subseteq (\sqcup\!\sqcup_T)^{\geq 2}(B_T(L))$. Thus, $v \in B_T(L) \cap (\sqcup\!\sqcup_T)^{\geq 2}(B_T(L))$, a contradiction. ■

Lemma 8.9.3 was established in a slightly weaker form by Ito and Silva [80, Prop. 3.2] for the case $T = (0+1)^*$. Hsiao *et al.* [69, Prop. 22] note the result for $T = 0^*1^* + 1^*0^*$.

The converse is given in a more general form as follows:

**Lemma 8.9.4** *Let $T \subseteq \{0,1\}^*$ be a set of trajectories such that $\pi(T)$ is sdl-preserving. Let $L \in \mathcal{P}_T(L)$. Then $L \cup \{\epsilon\}$ is del-$\pi(T)$ closed.*

**Proof.** As $L \in \mathcal{P}_T(L)$, $L \rightsquigarrow_{\pi(T)} L \subseteq \{\epsilon\}$, by (6.8). Consider that

$$\begin{aligned} & L \cup \{\epsilon\} \rightsquigarrow_{\pi(T)} L \cup \{\epsilon\} \\ = \; & (L \rightsquigarrow_{\pi(T)} L) \cup (\{\epsilon\} \rightsquigarrow_{\pi(T)} L \cup \{\epsilon\}) \cup (L \rightsquigarrow_{\pi(T)} \{\epsilon\}) \\ \subseteq \; & \{\epsilon\} \cup \{\epsilon\} \cup L = L \cup \{\epsilon\}. \end{aligned}$$

Therefore, $L \cup \{\epsilon\}$ is del-$\pi(T)$ closed, by Corollary 8.7.5. ■

## 8.10 $T$-Primitive Words

A word $w \in \Sigma^*$ is said to be *primitive* if $w = u^k$ implies $u = w$ and $k = 1$. For instance, the words $aab$ and $ababbb$ are primitive, while $aabaab = (aab)^2$ is not. The concept of primitive words is one of the most studied in formal language theory, and poses one of the most well-known of all open problems in formal language theory concerning the complexity of the set of all primitive words. The concept of a primitive word has been extended from concatenation to insertion and shuffle by Kari and Thierrin [118] and to arbitrary word operations by Hsiao *et al.* [69]. In this section, we consider primitivity with respect to a given shuffle on trajectories operation. We show that under our general definition of $(\sqcup\!\sqcup_T)^*$, every word has a $T$-primitive root. We also consider analogues of the Lyndon-Schützenberger Theorems for shuffle on trajectories.

### 8.10.1 $T$-Primitivity and $T$-roots

In this section, we consider primitive words with respect to a set of trajectories. Our definition will be with respect to our definition of iteration, which will alleviate certain restrictions which were imposed by Hsiao *et al.* [69].

Let $T \subseteq \{0, 1\}^*$. Given a word $w \in \Sigma^*$, we say that $w$ is *$T$-primitive* if $w \in (\sqcup\!\sqcup_T)^n(u)$ implies $u = w$ and $n = 1$. Let $Q_T(\Sigma)$ be the set of all $T$-primitive words over an alphabet $\Sigma$. For all $T \subseteq \{0, 1\}^*$ and $w \in \Sigma^*$, let

$$\sqrt[T]{w} = \{u \in Q_T(\Sigma) \ : \ w \in (\sqcup\!\sqcup_T)^+(u)\}.$$

We call $\sqrt[T]{w}$ the set of *$T$-primitive roots* of $w$.

It is well known (see, e.g, Lothaire [140, Prop. 1.3.1]) that every non-empty word has a unique primitive (i.e., $T$-primitive for $T = 0^*1^*$) root, that is, for $T = 0^*1^*$, $|\sqrt[T]{w}| = 1$ for all $w \in \Sigma^+$. Kari and Thierrin [118] note that for $T = 0^*1^*0^*$, this does not hold, as, e.g., $babb, bbab \in \sqrt[T]{(b^2a)^{n+1}b^{n+1}}$ for all $n \geq 1$. However, we now note that for all $T$, every non-empty word has at least one $T$-primitive root. We will require the following lemma:

**Lemma 8.10.1** *Let $u \in \Sigma^*$. Then $(\sqcup\!\sqcup_T)^+((\sqcup\!\sqcup_T)^+(u)) \subseteq (\sqcup\!\sqcup_T)^+(u)$.*

**Proof.** The proof follows by the fact that $(\sqcup\!\sqcup_T)^+(u)$ is shuffle-$T$ closed, by Theorem 8.6.5. ■

**Theorem 8.10.2** *Let $T \subseteq \{0, 1\}^*$. Let $w \in \Sigma^+$. Then $|\sqrt[T]{w}| \geq 1$.*

**Proof.** Let $w \in \Sigma^+$ be arbitrary. If $w \in Q_T(\Sigma)$, then we are done, as $w \in \sqrt[T]{w}$. Thus, assume that $w$ is not $T$-primitive.

Let $w \in (\sqcup\!\sqcup_T)^i(u)$ for some $u \in \Sigma^*$ and $i \geq 2$. If $u$ is $T$-primitive, then we are done, as $u \in \sqrt[T]{w}$.

Note that $|u| < |w|$ as $i \geq 2$. Now, if $u$ is not $T$-primitive, then $u \in (\sqcup\!\sqcup_T)^j(v)$ for some $v \in \Sigma^*$ and $j \geq 2$. Note that $w \in (\sqcup\!\sqcup_T)^+((\sqcup\!\sqcup_T)^+(v)) \subseteq (\sqcup\!\sqcup_T)^+(v)$. Thus, if $v$ is $T$-primitive, we are done.

Otherwise, as $|v| < |u| < |w|$, we note that as we continue this process, we eventually reach a word $x$ such that $w \in (\sqcup\!\sqcup_T)^+(x)$ and $x$ is $T$-primitive. This completes the proof. ■

### 8.10.2 Freeness and Uniqueness of $T$-Primitive Roots

We now turn to uniqueness of $T$-primitive roots. We will require the notion of a free semigroup, see, e.g., Shyr [184]. Recall that a semigroup is a set $S$ equipped with an associative binary operation; it does not necessarily have an identity element. Let $M$ be a semigroup. A non-empty subset $B \subseteq M$ is said to be a *base* for $M$ if and only if for all $u_1, u_2, \ldots, u_n, v_1, v_2, \ldots, v_m \in B$, the equality

$$u_1 u_2 \cdots u_n = v_1 v_2 \cdots v_m$$

implies that $n = m$ and $u_i = v_i$ for all $1 \leq i \leq n$. Note that $\Sigma$ is a base for $\Sigma^+$ as a semigroup under concatenation. A semigroup $M$ is said to be *free* if there exists some base $B$ such that $B^* = M$ (it can be easily shown that such a base must be unique). Levi [132] gives conditions on a semigroup being free. Let $T$ be an associative, deterministic set of trajectories. We say that $T$ is *free* if the semigroup $M = (\Sigma^+, \sqcup\!\sqcup_T)$ is free[2].

As we will be dealing exclusively with associative sets of trajectories, we will use the operation $(\sqcup\!\sqcup_T)^+_X$, which we gave in (8.5).

We first note that, besides concatenation, there exist non-trivial operations which are free. For instance, the operation of balanced insertion, given by $T = \{0^k 1^{2j} 0^k \ : \ k, j \geq 0\}$ is both deterministic and associative (see Mateescu *et al.* [147]) and is free, as is easily verified using Levi's conditions.

We now give an open problem concerning freeness:

**Open Problem 8.10.3** *Given $T$ regular (or context-free), is it decidable whether $T$ is free?*

It does not seem that Levi's conditions are helpful in this regard. Further, consider the following test: let $L = \Sigma^+ - (\Sigma^+ \sqcup\!\sqcup_T \Sigma^+)$. Test if $L$ is a $*$-T-code (i.e., every word in $(\sqcup\!\sqcup_T)^+(L)$ is uniquely generated). Then $T$ is free if and only if $L$ is a $*$-T-code.

[2] We could also easily frame the current discussion in terms of free monoids (i.e., semigroups with identities), and say that $T$ is free if the monoid $M = (\Sigma^*, \sqcup\!\sqcup_T, \epsilon)$ is free. However, as noted by Mateescu *et al.* [147, Rem. 4.7], we only need to require that $0^* + 1^* \subseteq T$ to ensure that $\epsilon$ is the identity element for $M$.

Since $T$ is regular implies that $L$ is regular, we expect that this would a plausible test for the freeness of $T$, since it is decidable whether a regular language is a $*$-code. However, the well-known algorithm to determine if a regular language is a $*$-code (e.g., Berstel and Perrin [18, Thm. I.3.1]) relies on the fact that $\Sigma^*$ is a free monoid under concatenation. Thus, it does not seem immediately possible to generalize this proof to other $T \subseteq \{0, 1\}^*$.

Levi's conditions are occasionally referred to as Levi's Lemma (see, e.g., Allouche and Shallit [3, Sect. 1.4]), which can be stated as follows for our purposes:

**Lemma 8.10.4** *Let $T$ be deterministic, associative and free. Then for all $u, v, x, y \in \Sigma^*$, $u \sqcup\!\sqcup_T v = x \sqcup\!\sqcup_T y$ implies that there exists $z \in \Sigma^*$ such that either $u = x \sqcup\!\sqcup_T z$ or $x = u \sqcup\!\sqcup_T z$.*

Levi's lemma (with $T = 0^*1^*$) is necessary for two of the most fundamental and elegant theorems in formal language theory and combinatorics on words, the Lyndon-Schützenberger Theorems (see, e.g., Allouche and Shallit [3, Sect. 1.4]):

**Theorem 8.10.5** *Let $x, z \in \Sigma^+$, $y \in \Sigma^*$. Then the following are equivalent:*

*(a) $xy = yz$;*

*(b) there exist $u, v \in \Sigma^*$, $e \geq 0$ such that $x = uv$, $z = vu$ and $y = (uv)^e u$.*

**Theorem 8.10.6** *Let $x, y \in \Sigma^+$. Then the following three conditions are equivalent:*

*(a) $xy = yx$;*

*(b) there exist integers $i, j > 0$ such that $x^i = y^j$;*

*(c) there exist $z \in \Sigma^+$ and integers $k, \ell > 0$ such that $x = z^k$ and $y = z^\ell$.*

We now show that freeness is the essential property in proving a generalization of the Lyndon-Schützenberger Theorems for $\sqcup\!\sqcup_T$.

The additional condition necessary to generalize the first Lyndon-Schützenberger Theorem is the possession of a unit element (see Mateescu *et al.* [147, Sect. 4.4]), which is the condition that $0^* + 1^* \subseteq T$. Thus, if $T$ has a unit element, $x \in (x \sqcup\!\sqcup_T \epsilon) \cap (\epsilon \sqcup\!\sqcup_T x)$ for all $x \in \Sigma^*$.

**Theorem 8.10.7** *Let $T$ be an associative, deterministic, free set of trajectories with a unit element. Let $x, z \in \Sigma^+$ and $y \in \Sigma^*$ be such that $x \sqcup\!\sqcup_T y, y \sqcup\!\sqcup_T z \neq \emptyset$. Then the following are equivalent:*

*(a)* $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T z$;

*(b) there exist $u, v \in \Sigma^*$, $e \geq 0$ such that*

$$\begin{aligned} x &= u \sqcup\!\sqcup_T v, \\ z &= v \sqcup\!\sqcup_T u, \\ y &= (\sqcup\!\sqcup_T)_X^e (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u. \end{aligned}$$

**Proof.** ((a) $\Rightarrow$ (b)): Let $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T z$. The proof is by induction on $|y|$. The base case for $y \in \Sigma^*$ is $|y| = 0$. In this case, $x = x \sqcup\!\sqcup_T \epsilon$ and $z = \epsilon \sqcup\!\sqcup_T z$, as $x \sqcup\!\sqcup_T \epsilon, \epsilon \sqcup\!\sqcup_T z \neq \emptyset$, by assumption. We conclude that $x = z$ and (b) holds with $u = \epsilon$, $v = x = z$ and $e = 0$.

Assume the result holds for all words $y'$ with $0 \leq |y'| < |y|$. Let $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T z$. By Lemma 8.10.4, there exists $w \in \Sigma^*$ such that either $x = y \sqcup\!\sqcup_T w$ and $z = w \sqcup\!\sqcup_T y$ or $y = x \sqcup\!\sqcup_T w$ and $y = w \sqcup\!\sqcup_T z$.

In the first case, let $u = y$, $v = w$ and $e = 0$. Then $y = u = (\sqcup\!\sqcup_T)_X^e (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u$, as $T$ is ST-strict. Further, $x = u \sqcup\!\sqcup_T v$ and $z = v \sqcup\!\sqcup_T u$.

In the second case, we have that $x \sqcup\!\sqcup_T w = w \sqcup\!\sqcup_T z$. Note that $|w| = |y| - |z| < |y|$ as $z \in \Sigma^+$. Therefore, by induction, there exist $u, v \in \Sigma^*$ and $e \geq 0$ such that $x = u \sqcup\!\sqcup_T v$, $z = v \sqcup\!\sqcup_T u$ and $w = (\sqcup\!\sqcup_T)_X^e (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u$. Note that

$$y = x \sqcup\!\sqcup_T w = (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^e (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u = (\sqcup\!\sqcup_T)_X^{e+1} (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u,$$

by the associativity of $T$. Thus, (b) is satisfied.

((b) $\Rightarrow$ (a)): Note that

$$\begin{aligned}
x \sqcup\!\sqcup_T y &= (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^e_X (u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T u \\
&= u \sqcup\!\sqcup_T v \sqcup\!\sqcup_T \underbrace{(u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T \cdots \sqcup\!\sqcup_T (u \sqcup\!\sqcup_T v)}_{e \text{ times.}} \sqcup\!\sqcup_T u \\
&= \underbrace{(u \sqcup\!\sqcup_T v) \sqcup\!\sqcup_T \cdots \sqcup\!\sqcup_T (u \sqcup\!\sqcup_T v)}_{e \text{ times.}} \sqcup\!\sqcup_T u \sqcup\!\sqcup_T v \sqcup\!\sqcup_T u \\
&= y \sqcup\!\sqcup_T z.
\end{aligned}$$

This proves the desired equality. ■

Call a set $T$ of trajectories *concatenation-like* if $T$ is deterministic, associative, free, and satisfies the following property, which we call *power-enabling*: for all $n, k \geq 0$, $(kn, n) \in \Psi(T) \Rightarrow ((k+1)n, n) \in \Psi(T)$. Note that balanced insertion is power-enabling and hence concatenation-like.

We will require the following proposition, which is easily proven by induction:

**Proposition 8.10.8** *Let $T$ be associative. Then for all $k, \ell > 0$, and all $z \in \Sigma^+$,*

$$(\sqcup\!\sqcup_T)^k_X((\sqcup\!\sqcup_T)^\ell_X(z)) = (\sqcup\!\sqcup_T)^{k\ell}_X(z).$$

We may now state and prove our second generalized Lyndon-Schützenberger Theorem:

**Theorem 8.10.9** *Let $T$ be concatenation-like. Let $x, y \in \Sigma^+$ be words such that $x \sqcup\!\sqcup_T y$ and $y \sqcup\!\sqcup_T x$ are non-empty. Then the following three conditions are equivalent:*

*(a) $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T x$;*

*(b) there exist integers $i, j > 0$ such that $\emptyset \neq (\sqcup\!\sqcup_T)^i_X(x) = (\sqcup\!\sqcup_T)^j_X(y)$;*

*(c) there exist $z \in \Sigma^+$ and $k, \ell > 0$ such that $x = (\sqcup\!\sqcup_T)^k_X(z)$ and $y = (\sqcup\!\sqcup_T)^\ell_X(z)$.*

**Proof.** ((a) $\Rightarrow$ (c)): Assume that $x \neq y$ and $\emptyset \neq x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T x$. We show the result by induction on $|xy|$. As $x, y \in \Sigma^+$, the base case is $|xy| = 2$. Thus, $x, y \in \Sigma$. Thus, $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T x$, and $T$ deterministic implies that $x = y$, contrary to our assumption.

Assume that the result holds for all $x, y \in \Sigma^+$ with $|xy| < n$. Let $|xy| = n$. Let $|x| \geq |y|$. As $x \sqcup\!\sqcup_T y = y \sqcup\!\sqcup_T x$, by Levi's property, there exists $w \in \Sigma^*$ such that $x = y \sqcup\!\sqcup_T w$. Note that

$y \sqcup_T x = x \sqcup_T y = (y \sqcup_T w) \sqcup_T y = y \sqcup_T (w \sqcup_T y)$ by the associativity of $T$. Thus, by the determinism of $T$, $x = w \sqcup_T y$ and $w \sqcup_T y = y \sqcup_T w$.

If $w = y$, then $x = y \sqcup_T y$, and (c) follows with $z = y$ and $k = 2$, $\ell = 1$. Thus, assume that $w \neq y$.

If $|w| = 0$, then $x = y \sqcup_T w$ implies that $x = y$. Again, this contradicts our assumptions on $x, y$. Thus, $|w| > 0$. Note that $|wy| = |x| < |xy|$. Thus, as $w \sqcup_T y = y \sqcup_T w$, by induction there exist $z \in \Sigma^+$, $k, \ell > 0$ such that $w = (\sqcup_T)_X^k(z)$ and $y = (\sqcup_T)_X^\ell(z)$. Thus,

$$x = (\sqcup_T)_X^\ell(z) \sqcup_T (\sqcup_T)_X^k(z).$$

Thus, $x \in (\sqcup_T)_X^+(z)$ by the closure of $(\sqcup_T)_X^+(z)$ under $\sqcup_T$. As $T$ is deterministic, we have that there is some $m > 0$ such that $x = (\sqcup_T)_X^m(z)$. Thus, (c) is satisfied.

((c) $\Rightarrow$ (b)): Let $z \in \Sigma^+$, $k, \ell > 0$ be such that $x = (\sqcup_T)_X^k(z)$ and $y = (\sqcup_T)_X^\ell(z)$. Using Proposition 8.10.8, we note that

$$(\sqcup_T)_X^\ell(x) = (\sqcup_T)_X^\ell((\sqcup_T)_X^k(z)) = (\sqcup_T)_X^{\ell k}(z) = (\sqcup_T)_X^k((\sqcup_T)_X^\ell(z)) = (\sqcup_T)_X^k(y).$$

By the fact that $(\sqcup_T)_X^k(z) = x$, $((k-1)|z|, |z|) \in \Psi(T)$. Thus, $((\ell \cdot k - 1)|z|, |z|) \in \Psi(T)$ and both $(\sqcup_T)_X^\ell(x)$ and $(\sqcup_T)_X^k(y)$ are non-empty. Thus, (b) is satisfied.

((b) $\Rightarrow$ (a)): Let $i, j > 0$ be such that $\emptyset \neq (\sqcup_T)_X^i(x) = (\sqcup_T)_X^j(y)$. Note that if $|x| = |y|$, then $|(\sqcup_T)_X^i(x)| = |(\sqcup_T)_X^j(y)|$ implies that $i = j$. Thus, by the determinism of $T$, $x = y$ and (a) holds.

Thus, without loss of generality, assume that $|x| > |y|$. Thus, by the associativity of $T$,

$$x \sqcup_T (\sqcup_T)_X^{i-1}(x) = (\sqcup_T)_X^i(x) = (\sqcup_T)_X^j(y) = y \sqcup_T (\sqcup_T)_X^{j-1}(y).$$

(We let $(\sqcup_T)_X^{j-1}(y) = \epsilon$ if $j = 1$, and similarly for $x$ and $i$.) Thus, by Levi's property, there exists some $w \in \Sigma^+$ such that $x = y \sqcup_T w$. Thus,

$$(\sqcup_T)_X^j(y) = (\sqcup_T)_X^i(x) = (\sqcup_T)_X^i(y \sqcup_T w).$$

Note that

$$(\sqcup\!\sqcup_T)^i_X(y \sqcup\!\sqcup_T w) = \underbrace{(y \sqcup\!\sqcup_T w) \sqcup\!\sqcup_T (y \sqcup\!\sqcup_T w) \sqcup\!\sqcup_T \cdots \sqcup\!\sqcup_T (y \sqcup\!\sqcup_T w)}_{i \text{ times}}.$$

By the determinism of $T$, $(\sqcup\!\sqcup_T)^j_X(y) = (\sqcup\!\sqcup_T)^i_X(y \sqcup\!\sqcup_T w)$ implies that

$$\begin{array}{rrcl} & (\sqcup\!\sqcup_T)^{j-1}_X(y) & = & w \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^{i-1}_X(y \sqcup\!\sqcup_T w) \\ \Rightarrow & (\sqcup\!\sqcup_T)^{j-1}_X(y) \sqcup\!\sqcup_T y & = & w \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^{i-1}_X(y \sqcup\!\sqcup_T w) \sqcup\!\sqcup_T y \\ \Rightarrow & (\sqcup\!\sqcup_T)^{j}_X(y) & = & (\sqcup\!\sqcup_T)^{i}_X(w \sqcup\!\sqcup_T y) \\ \Rightarrow & (\sqcup\!\sqcup_T)^{i}_X(y \sqcup\!\sqcup_T w) & = & (\sqcup\!\sqcup_T)^{i}_X(w \sqcup\!\sqcup_T y). \end{array}$$

By the determinism of $T$, $y \sqcup\!\sqcup_T w = w \sqcup\!\sqcup_T y$. Thus,

$$\begin{array}{rcl} x \sqcup\!\sqcup_T y & = & (y \sqcup\!\sqcup_T w) \sqcup\!\sqcup_T y \\ & = & y \sqcup\!\sqcup_T (w \sqcup\!\sqcup_T y) \\ & = & y \sqcup\!\sqcup_T x. \end{array}$$

Thus, (a) is satisfied, as $y \neq x$ and $y \sqcup\!\sqcup_T x = x \sqcup\!\sqcup_T y \neq \emptyset$. ■

Our main corollary of the generalized Lyndon-Schützenberger Theorems concerns the uniqueness of $T$-primitive roots:

**Corollary 8.10.10** *Let $T$ be concatenation-like. Then for all $u \in \Sigma^+$, $u$ has a unique $T$-primitive root.*

**Proof.** Let $u \in \Sigma^+$, and assume that $v_1, v_2 \in \sqrt[T]{u}$. Then as $T$ is concatenation-like, we have $u = (\sqcup\!\sqcup_T)^{i_1}(v_1)$ and $u = (\sqcup\!\sqcup_T)^{i_2}(v_2)$ for some $i_1, i_2 \geq 1$. Thus, there exist $j_1, j_2$ such that $u = (\sqcup\!\sqcup_T)^{j_1}_X(v_1)$ and $u = (\sqcup\!\sqcup_T)^{j_2}_X(v_2)$.

By Theorem 8.10.9, there exist $z \in \Sigma^+$ and $k_1, k_2 \geq 1$ such that $v_1 = (\sqcup\!\sqcup_T)^{k_1}_X(z)$ and $v_2 = (\sqcup\!\sqcup_T)^{k_2}_X(z)$. As $v_1, v_2 \in Q_T(\Sigma)$, we must have that $k_1 = k_2 = 1$ and $z = v_1 = v_2$. Thus, $|\sqrt[T]{u}| = 1$, as required. ■

Thus, we see that if $T$ is catenation-like, then each word has a unique $T$-primitive root. This applies, e.g., to balanced insertion.

## 8.11 Language Equations Revisited

In Chapter 7, we have studied language equations involving shuffle and deletion along trajectories. In this section, we revisit this topic to consider some equation forms which we did not consider in Chapter 7.

In particular, we note that all of language equations we have considered have the form

$$L = X \diamond Y$$

where $X, Y$ (and occasionally $\diamond$) may be unknown, but the language $L$ is always fixed. We recall that, in the terminology of, e.g., Leiss [131, Sect. 2.6], such equations are called *implicit language equations*. In this section, we consider *explicit language equations*, which we recall are of the form $X = \alpha$ where $X$ is unknown, and $\alpha$ is an equation involving constants, language operations (including $\sqcup\!\sqcup_T$) and unknowns.

Clearly, explicit language equations are of considerable interest; indeed, the fundamental notion of CFGs is equivalent to explicit systems of language equations of the form $X = \alpha$ where $\alpha$ is an expression involving catenation and union [10, Sect. 2]. Extensions of the context-free grammar formalism to capture larger classes of languages via grammars have also been studied, for example, conjunctive [157] and Boolean [159] grammars, both recently introduced by Okhotin. Explicit language equations are crucial to both these studies.

Consideration of language equations with unknowns on both sides of the equality sign must involve some caution, however, especially with regard to our interest thus far in answering questions such as "is it decidable whether this equation has a solution?" Consider the equation $X = L_1 \sqcup\!\sqcup_T L_2$, where $L_1, L_2$ are languages and $T$ is a set of trajectories. Clearly, this equation has a solution $X = L_1 \sqcup\!\sqcup_T L_2$. Note that this assumes nothing about the complexity of $L_1, L_2$ or $T$. Other equations possess trivial solutions; $X = X \sqcup\!\sqcup_T L$ has a solution $X = \emptyset$ regardless of $L$ and $T$.

Thus, in this section, our focus will shift somewhat from decidability, which is often trivial, to characterizations of solutions. Our results will be similar in spirit to Arden's Lemma (Arden [9], cf., Salomaa [173, Ch. 3]) which states that the equation $X = XR_1 + R_2$ has a unique solution $R_1^* R_2$ if $\epsilon \notin R_1$. Further, $R_1^* R_2$ is the least solution (by inclusion) of $X = XR_1 + R_2$, independent of $R_1$ (see, e.g., Conway [28, Thm. 3, p. 27]). We will see that such results can be extended to $\sqcup\!\sqcup_T$, under certain assumptions on $T$, and that related equations have similar characterizations.

### 8.11.1 Arden-like Equations

Our first consideration will be the following equation:

$$X = X \sqcup\!\sqcup_T L_1 + L_2,$$

where $X$ is unknown, and $L_1, L_2$ are arbitrary languages. In the case where $T = 0^* 1^*$, the characterization of the unique solution of this equation (under the assumption that $\epsilon \notin L_1$) is known as Arden's Lemma (however, the identity $L_2 L_1^* = L_2 L_1^* L_1 + L_2$ was previously known, see, e.g., Kleene [119, Eq. (9), p. 24]). We will require some conditions on $T$ in order for a similar result to hold.

**Theorem 8.11.1** *Let $T \subseteq \{0,1\}^*$ be left-preserving and associative. Let $L_1, L_2$ be arbitrary languages. The least solution to the equation*

$$X = X \sqcup\!\sqcup_T L_1 + L_2, \tag{8.20}$$

*is the language $L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^*(L_1)$.*

**Proof.** As $T$ is associative, it will suffice to establish that the language $L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^*(L_1)$ is the least solution to (8.20). Recall that $(\sqcup\!\sqcup_T)_X^*$ is defined by (8.4).

We first show that $L = L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^*(L_1)$ is a solution to (8.20). We establish the inclusion $L \sqcup\!\sqcup_T L_1 + L_2 \subseteq L$, by proving that

$$L_2 \subseteq L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^*(L_1) \tag{8.21}$$

and that

$$\left(L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^i(L_1)\right) \sqcup\!\sqcup_T L_1 \subseteq L_2 \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)_X^*(L_1) \tag{8.22}$$

for all $i \geq 0$. We note that (8.21) holds since $T$ is left-preserving and $\epsilon \in (\sqcup_T)^*_X(L_1)$. We now establish (8.22) for all $i \geq 0$. Note that for arbitrary $i \geq 0$,

$$\begin{aligned}(L_2 \sqcup_T (\sqcup_T)^i_X(L_1)) \sqcup_T L_1 &= L_2 \sqcup_T ((\sqcup_T)^i_X(L_1) \sqcup_T L_1)\\ &= L_2 \sqcup_T (\sqcup_T)^{i+1}_X(L_1)\\ &\subseteq L_2 \sqcup_T (\sqcup_T)^*_X(L_1).\end{aligned}$$

We now establish the inclusion $L \sqcup_T L_1 + L_2 \supseteq L$. Let $x \in L = L_2 \sqcup_T (\sqcup_T)^*_X(L_1)$. Then there exists $i \geq 0$ such that $x \in L_2 \sqcup_T (\sqcup_T)^i_X(L_1)$. For $i = 0$, $x \in L_2 \sqcup_T \{\epsilon\} \subseteq L_2$. Thus, $x \in L_2 \subseteq L \sqcup_T L_1 + L_2$. Let $i > 0$. Then

$$\begin{aligned}x &\in L_2 \sqcup_T (\sqcup_T)^i_X(L_1)\\ &= L_2 \sqcup_T ((\sqcup_T)^{i-1}_X(L_1) \sqcup_T L_1)\\ &= (L_2 \sqcup_T (\sqcup_T)^{i-1}_X(L_1)) \sqcup_T L_1\\ &\subseteq (L_2 \sqcup_T (\sqcup_T)^*_X(L_1)) \sqcup_T L_1 + L_2.\end{aligned}$$

Thus, $L$ is a solution to (8.20). We now show that it is the least solution to this equation. Let $X_0$ be the least solution to (8.20). We show that $X_0 \supseteq L$. First, we note that

$$X_0 = X_0 \sqcup_T L_1 + L_2 \supseteq L_2 = L_2 \sqcup_T (\sqcup_T)^0_X(L_1).$$

Let $i \geq 0$. Assume that $X_0 \supseteq L_2 \sqcup_T (\sqcup_T)^i_X(L_1)$. Then we have that

$$\begin{aligned}X_0 \supseteq X_0 \sqcup_T L_1 &\supseteq (L_2 \sqcup_T (\sqcup_T)^i_X(L_1)) \sqcup_T L_1\\ &\supseteq L_2 \sqcup_T ((\sqcup_T)^i_X(L_1) \sqcup_T L_1)\\ &\supseteq L_2 \sqcup_T (\sqcup_T)^{i+1}_X(L_1).\end{aligned}$$

Thus, $X_0 \supseteq L_2 \sqcup_T (\sqcup_T)^*_X(L_1) = L$. This completes the proof. ■

We can also give the following symmetric result.

**Theorem 8.11.2** *Let $T \subseteq \{0,1\}^*$ be right-preserving and associative. Let $L_1, L_2$ be arbitrary languages. Then the least solution to the equation*

$$X = L_1 \sqcup\!\sqcup_T X + L_2, \tag{8.23}$$

*is the language $(\sqcup\!\sqcup_T)^*(L_1) \sqcup\!\sqcup_T L_2$.*

Further results in this area are likely. For instance, Salomaa considers systems of equations of the form

$$X_i = \sum_{j=1}^{n} X_j L_{j,i} + R_i$$

for $1 \le i \le n$, and investigates their solutions [173, Ch. 3, Sect. 2]. Systems of this form with $X_j \sqcup\!\sqcup_T L_{j,i}$ are clear generalizations (for a single fixed $T \subseteq \{0,1\}^*$), and results can likely be obtained in the same manner as Theorems 8.11.1 and 8.11.2.

### 8.11.2 A Language Equation for $(\sqcup\!\sqcup_T)^+(L)$

Due to the conditions placed on $T$ in the previous section, the following question seems natural: given arbitrary $T$ and $L$, can we find a language equation for which $(\sqcup\!\sqcup_T)^+(L)$ is the minimal solution? We give this equation in the following theorem:

**Theorem 8.11.3** *Let $T \subseteq \{0,1\}^*$. For any language $L \subseteq \Sigma^*$, the least solution to the equation*

$$X = X \sqcup\!\sqcup_T X + L \tag{8.24}$$

*is $X = (\sqcup\!\sqcup_T)^+(L)$.*

**Proof.** We first show that $(\sqcup\!\sqcup_T)^+(L)$ is a solution of (8.24). Consider that

$$\begin{aligned}(\sqcup\!\sqcup_T)^+(L) \sqcup\!\sqcup_T (\sqcup\!\sqcup_T)^+(L) + L &\subseteq (\sqcup\!\sqcup_T)^+(L) + L \\ &\subseteq (\sqcup\!\sqcup_T)^+(L).\end{aligned}$$

The last inclusion is by definition of $(\sqcup\!\sqcup_T)^+(L)$ and the first is by Proposition 8.6.2 and Theorem 8.6.5. To prove the reverse inclusion, let $x \in (\sqcup\!\sqcup_T)^+(L)$. Then there exists $i \ge 1$ such that $x \in$

$(\sqcup_T)^i(L)$. If $i = 1$, then $x \in L \subseteq (\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L$, and the inclusion holds. Assume then that $i > 1$ and for all $y \in (\sqcup_T)^j(L)$ with $j < i$, $y \in (\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L$. Let $x \in (\sqcup_T)^i(L)$. Then by definition, $x \in (\sqcup_T)^{i-1}(L)$ or $x \in (\sqcup_T)^{i-1}(L) \sqcup_T (\sqcup_T)^{i-1}(L)$. In the first case, the result holds by induction. Otherwise,

$$\begin{aligned} x &\in (\sqcup_T)^{i-1}(L) \sqcup_T (\sqcup_T)^{i-1}(L) \\ &\subseteq ((\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L) \sqcup_T ((\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L) \\ &\subseteq ((\sqcup_T)^+(L) + L) \sqcup_T ((\sqcup_T)^+(L) + L) \\ &\subseteq (\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L. \end{aligned}$$

Here, the first inclusion is by induction, the second is by Proposition 8.6.2 and Theorem 8.6.5, and the third is by the distributivity of $(\sqcup_T)^+$ over union. Thus,

$$(\sqcup_T)^+(L) = (\sqcup_T)^+(L) \sqcup_T (\sqcup_T)^+(L) + L.$$

Let $X_0 \subseteq \Sigma^*$ be the least solution of (8.24). As $(\sqcup_T)^+(L)$ is a solution of (8.24), $X_0 \subseteq (\sqcup_T)^+(L)$. It remains to show the reverse inclusion, which is again accomplished by induction, by showing that $(\sqcup_T)^i(L) \subseteq X_0$ for all $i \geq 1$.

For $i = 1$, $L \subseteq L + X_0 \sqcup_T X_0 = X_0$. Let $i > 1$ and assume that $(\sqcup_T)^{i-1}(L) \subseteq X_0$. Then note that

$$\begin{aligned} X_0 &= X_0 \sqcup_T X_0 + L \\ &\supseteq X_0 \sqcup_T X_0 \\ &\supseteq (\sqcup_T)^{i-1}(L) \sqcup_T (\sqcup_T)^{i-1}(L). \end{aligned}$$

Thus $X_0 \supseteq (\sqcup_T)^{i-1}(L)$ and $X_0 \supseteq (\sqcup_T)^{i-1}(L) \sqcup_T (\sqcup_T)^{i-1}(L)$, i.e., $X_0 \supseteq (\sqcup_T)^{i-1}(L) + (\sqcup_T)^{i-1}(L) \sqcup_T (\sqcup_T)^{i-1}(L) = (\sqcup_T)^i(L)$. This establishes the inclusion. Thus, $X_0 = (\sqcup_T)^+(L)$, establishing the result. ■

## 8.12 Conclusions

In this chapter, we have considered the iteration of shuffle on trajectory and deletion on trajectory operations. Our work generalizes previous work by Ito *et al.* [78, 79] and Kari and Thierrin [115] on particular shuffle and deletion along trajectories operations. Our work is also similar to that of Hsiao *et al.* [69] on iteration of arbitrary word operations. However, by considering a more general definition of iterated shuffle on trajectories, we are able to overcome some of the conditions imposed by Hsiao *et al.* in their study of iterated binary word operations.

After considering iterated shuffle and deletion along trajectories and the closure of languages under shuffle and deletion along trajectories, we have investigated the notions of shuffle base, extended shuffle base, primitivity and the Lyndon-Schützenberger Theorems, all considered in trajectory-based contexts. These notions generalize well to shuffle and deletion along trajectories, and many key concepts hold. We note that in the context of investigating the Lyndon-Schützenberger Theorems, we raise the problem of when $ш_T$ defines a free operation. This fundamental problem remains open.

Finally, we have returned to the topic of language equations. For explicit equations, the problem of decidability of the existence of solutions becomes trivial, and we have instead focused on characterizing least solutions to language equations. We have found explicit least solutions for two key explicit language equations involving $ш_T$. One generalizes a well-known language equation for Kleene closure which has been studied for fifty years. The second is a general language equation formulated so that its least solution is $(ш_T)^+(L)$.