On the Separability Problem of String Constraints

Parosh Aziz Abdulla  
Uppsala University, Sweden  
parosh@it.uu.se

Mohamed Faouzi Atig  
Uppsala University, Sweden  
mohamed_faouzi.atig@it.uu.se

Vrunda Dave  
IIT Bombay, India  
vrunda@cse.iitb.ac.in

Shankara Narayanan Krishna  
IIT Bombay, India  
krishnas@cse.iitb.ac.in

Abstract

We address the separability problem for straight-line string constraints. The separability problem for languages of a class C by a class S asks: given two languages A and B in C, does there exist a language I in S separating A and B (i.e., I is a superset of A and disjoint from B)? The separability of string constraints is the same as the fundamental problem of interpolation for string constraints. We first show that regular separability of straight line string constraints is undecidable. Our second result is the decidability of the separability problem for straight-line string constraints by piece-wise testable languages, though the precise complexity is open. In our third result, we consider the positive fragment of piece-wise testable languages as a separator, and obtain an EXPSPACE algorithm for the separability of a useful class of straight-line string constraints, and a PSPACE-hardness result.

2012 ACM Subject Classification  
Security and privacy → Logic and verification; Theory of computation → Verification by model checking

Keywords and phrases  
string constraints, separability, interpolants

Digital Object Identifier  
10.4230/LIPIcs...
On the Separability Problem of String Constraints

subclass, that has been studied extensively, is that of straight-line (SL) string constraints (e.g., [26, 14, 32, 26, 12]). The SL fragment was introduced by Barceló and Lin in [32]. Roughly, an SL constraint models the feasibility of a path of a string-manipulating program that can be generated by symbolic execution. The satisfiability of the SL fragment was shown to be EXPSPACE-complete in [32] and forms the basis of many of the tools above [26, 12].

In this paper, we focus on the fundamental problem of interpolation/separability for the SL fragment of string constraints. An interpolant for a pair of formulas \( A, B \) is a formula over their common vocabulary that is implied by \( A \) and is inconsistent with \( B \). The Craig-Lyndon interpolation technique is very well-known in mathematical logic. McMillan [33] in his pioneering work, has also recognized interpolation as an efficient method for automated construction of abstractions of systems. Interpolation based algorithms have been developed for a number of problems in program verification [33, 34, 35].

Interpolation procedures have been implemented by many solvers for the theories most commonly used in program verification like linear arithmetic, uninterpreted functions with equality and some combination of such theories. In most of these algorithms, the interpolants were simple. The interpolation technique can also be used to check the unsatisfiability. In fact, the existence of an interpolant for formulas \( A \) and \( B \) implies the unsatisfiability of \( A \land B \).

The notion of separators in formal language theory is the counterpart of interpolants in logic. The separability problem for languages of a class \( \mathcal{C} \) by a class \( \mathcal{S} \) asks: given two languages \( I, E \in \mathcal{C} \), does there exist a language \( S \in \mathcal{S} \) separating \( I \) and \( E \)? That is, \( I \subseteq S \) and \( S \cap E = \emptyset \). The language \( S \) is called the separator of \( I, E \). Separability is a classical problem of fundamental interest in theoretical computer science, and has recently received a lot of attention. For instance, regular separability has been studied for one-counter automata [17], Parikh automata [15], and well-structured transition systems [18]. In the following, we use the terms interpolant or separator of two SL string constraints to mean the same thing, since the solutions of a string constraint can be interpreted as a language.

In this paper, we first show that any string constraint \( \phi \) can be written as the conjunction of two SL string constraints \( A \) and \( B \). Therefore, the interpolation problem for the pair \( A \) and \( B \) can be used to check the unsatisfiability of the string constraint \( \phi \). (Recall that the satisfiability problem for general string constraints is undecidable [14].)

Then, we consider the regular separability problem for SL string constraints. We show that this problem is undecidable (Theorem 2) by a reduction from the halting problem of Turing Machines. The main technical difficulty here is to ensure that the encoding of a sequence of configurations of a Turing machine results in SL string constraints.

Due to this undecidability, we focus on the separability problem of SL string constraints by piece-wise testable languages (PTL). A PTL is a finite Boolean combination of special regular languages called piece languages of the form \( \Sigma^*a_1\Sigma^*a_2\ldots\Sigma^*a_n\Sigma^* \), where all \( a_j \in \Sigma \). PTL is a very natural and well-studied class of languages in the context of the separability problem (e.g. [37, 19, 20]). Furthermore, among the various separator classes considered in the literature, the class of piecewise testable languages (PTL) seems to be the most tractable: PTL-separability of regular languages is in PTIME [37, 19]. To decide the PTL-separability of SL string constraints, we first encode the solutions of an SL string constraint as the language of an Ordered Multi-Pushdown Automaton (OMPA) (Section 4.1). Then, we show that the PTL-separability of SL constraints can be reduced to the PTL-separability of OMPAs. To show the decidability of the latter problem, we first prove that the language of an OMPA: (1) is a full trio [21] and (2) has a semilinear Parikh image. Using (1), we obtain the equivalence of the PTL separability problem and the diagonal problem for OMPAs from [20], where the
equivalence has been shown to hold for full trios. Next, the decidability of PTL-separability problem for OMPAs is obtained from the decidability of the diagonal problem for OMPAs: the latter is obtained using (2) and [20] where the decidability of the diagonal problem has been shown for languages having a semilinear Parikh image. As a corollary of these results, we obtain the decidability of the PTL-separability problem for SL string constraints and OMPAs; however the exact complexity is still an open question. In fact, it is an open problem in the case of OMPAs with one stack (i.e., Context-Free Languages (CFLs)) [20].

Given the complexity question, we propose the class of positive piecewise testable languages (PosPTL) as separators. PosPTL is obtained as a negation-free Boolean combination of piece languages. As a first result (Theorem [12]) we show that deciding PosPTL-separability for any language class has a very elegant proof: it suffices to check if the upward (downward) closure of one of the languages is disjoint from the other language. Using this result, we prove the PSPACE-completeness of the PosPTL-separability for CFLs, thereby progressing on the complexity front with respect to a problem which is open in the case of PTL-separability for CFLs. Then, we focus on a class of SL string constraints where the variables used in outputs of the transducers are independent of each other. This class contains SL string constraints with functional transducers (computing partial functions, by associating at most one output with each input). We prove the decidability and EXPSPACE membership for the PosPTL-separability of this class by first encoding the solutions of string constraints as outputs of two way transducers (2NFT), and then proving the decidability of PosPTL-separability for 2NFT.

Due to lack of space, missing proofs of the technical results can be found in the appendix.

Related work. The satisfiability problem for string constraints is an active research area and there is a lot of progress in the last decade (e.g., [38] [20] [22] [12] [14] [5] [22] [23] [4] [47]). An interpolation based semi-decision procedure for string constraints has been proposed in [4]. As far as we know, this is the first time the separability problem has been studied in the context of string constraints.

2 Preliminaries

Notations. Let [i, j] denote the set {i, ..., j} for i, j ∈ N. Let Σ be a finite alphabet. Σ* denotes the set of all finite words over Σ and Σ+ denotes Σ* \ {ε} where ε is the empty word. We denote Σ ∪ {ε} by Σ+. Let u ∈ Σ+. We use uR to denote the reverse of u. The length of the word u is denoted |u| and the ith symbol of u by u[i]. Given two words u ∈ Σ* and v ∈ Σ*, we say that u is a subword of v (denoted u ≤ v) if there is a mapping h : [1, |u|] → [1, |v|] such that (1) u[i] = v[h(i)] for all i ∈ [1, |u|], and (2) h(i) < h(j) for all i < j.

(Multi-tape)-Automata. A Finite State Automaton (FSA) over an alphabet Σ is a tuple A = (Q, Σ, δ, I, F), where Q is a finite set of states, δ ⊆ Q × Σ × Q is a set of transitions, and I ⊆ Q (resp. F ⊆ Q) are the initial (resp. accepting) states. A accepts a word w iff there is a sequence q0a1q1a2...aqn such that (qi−1, ai, qi) ∈ δ for all 1 ≤ i ≤ n, q0 ∈ I, qn ∈ F, and w = a1 · · · an. The language of A, denoted L(A), is the set all accepted words.

Given n ∈ N, a n-tape automaton T is an automaton over the alphabet (Σ*)n. It recognizes the relation R(T) ⊆ (Σ*)n that contains the n-tuple of words (w1, w2, . . . , wn) for which there is a word (a1(1), a2(1), . . . , an(1)) · · · (a1(m), a2(m), . . . , am(m)) ∈ L(T) with wi = a(i,1) · · · a(i,m) for all i ∈ {1, . . . , n}. A transducer is a 2-tape automaton.

Well-quasi orders. Given a (possibly infinite set) C, a quasi-order on C is a reflexive and transitive relation ⊆⊆ C × C. An infinite sequence c1, c2, ... in C is said to be saturating if there exists indices i < j s.t. ci ⊆ cj. A quasi-order ⊆ is said to be a well-quasi order
On the Separability Problem of String Constraints

(wqo) on C if every infinite sequence in C is saturating. Observe that the subword ordering \( \preceq \) between words \( u,v \) over a finite alphabet \( \Sigma \) is well-known to be a wqo on \( \Sigma^* \).

Upward and Downward Closure. Given a wqo \( \sqsubseteq \) on a set \( C \), a set \( U \subseteq C \) is said to be upward closed if for every \( a \in U \) and \( b \in C \), with \( a \sqsubseteq b \), we have \( b \in U \). The upward closure of a set \( U \subseteq C \) is defined as \( U^\uparrow = \{ b \in C \mid \exists a \in U, a \sqsubseteq b \} \). It is known that every upward closed set \( U \) can be characterized by a finite minor. A minor \( M \subseteq U \) is s.t. (i) for each \( a \in U \), there is a \( b \in M \) s.t. \( b \sqsubseteq a \), and (ii) for all \( a,b \in M \) s.t. \( a \preceq b \), we have \( a = b \). For an upward closed set \( U \), let \( \min \) be the function that returns the minor of \( U \). Downward closures are defined analogously. The downward closure of a set \( D \subseteq C \) is defined as \( D^\downarrow = \{ b \in C \mid \exists a \in D, b \sqsubseteq a \} \). The notion of subword relation and thus upward and downward closures naturally extends to \( n \)-tuples of words. The subword relation here is component-wise i.e. \( (u_1,\ldots,u_n) \preceq_n (v_1,\ldots,v_n) \) iff \( u_i \preceq v_i \) for all \( i \in [1,n] \).

String Constraints. An atomic string constraint \( \varphi \) over an alphabet \( \Sigma \) and a set of string variables \( X \) is either: (1) a membership constraint of the form \( x \in L(A) \) where \( x \in X \) and \( A \) is a FSA (i.e., the evaluation of \( x \) is in the language of a FSA \( A \) over \( \Sigma \)), or (2) a relational constraint of the form \((t',t) \in R(T)\) where \( t \) and \( t' \) are string terms (i.e., concatenation of variables in \( X \)) and \( T \) is a transducer over \( \Sigma \), and \( t \) and \( t' \) are related by a relation recognised by the transducer \( T \). \((t',t) \in R(T)\) can also be written as \( t' = T(t) \), that is, \( T \) produces \( t' \) as the output on input \( t \). For a given term \( t \), \(|t|\) denotes the number of variables appearing in \( t \).

A string constraint \( \Psi \) is a conjunction of atomic string constraints. We define the semantics of string constraints using a mapping \( \eta \), called evaluation, that assigns for each variable a word over \( \Sigma \). The evaluation \( \eta \) can be extended in the straightforward manner to string terms as follows: \( \eta(t_1 \cdot t_2) = \eta(t_1) \cdot \eta(t_2) \). We extend also \( \eta \) to atomic constraints as follows: (1) \( \eta(x \in L(A)) = \top \) iff \( \eta(x) \in L(A) \), and (2) \( \eta((t,t') \in R(T)) = \top \) iff \( \eta(t),\eta(t') \in R(T) \). The truth value of \( \Psi \) for an evaluation \( \eta \) is defined in the standard manner. If \( \eta(\Psi) = \top \) then \( \eta \) is a solution of \( \Psi \), written \( \eta \models \Psi \). The formula \( \Psi \) is satisfiable iff it has a solution.

A string constraint is said to be Straight Line [1] (SL) if it can be rewritten as \( \Psi' \wedge \bigwedge_{i=1}^k \varphi_i \) where \( \Psi' \) is a conjunction of membership constraints, and \( \varphi_1,\ldots,\varphi_k \) are relational constraints such that (1) there is a sequence of different string variables \( x_1,x_2,\ldots,x_n \) with \( n \geq k \), and (2) \( \varphi_i \) is of the form \((x_i,t_i) \in R(T_i)\) such that if a variable \( x_j \) is appearing in \( t_i \) then \( j > i \).

A string constraint in the SL form is called an SL formula. Observe that any string formula can be rewritten as a conjunction of two SL formulas (by using extra-variables).

\[ \triangleright \textbf{Lemma 1.} \text{ Given a string constraint } \Psi, \text{ it is possible to construct two SL string constraints } \Psi_1 \text{ and } \Psi_2 \text{ such that } \Psi \text{ is satisfiable iff } \Psi_1 \wedge \Psi_2 \text{ is satisfiable.} \]

Let \( \Psi \) be a string constraint and \( x_1,\ldots,x_n \) be the set of variables appearing in \( \Psi \). We use \( L(\Psi) \) to denote the language of \( \Psi \) which consists of the set of \( n \)-tuple of words \( (u_1,\ldots,u_n) \) such that there is an evaluation \( \eta \) with \( \eta(\Psi) = \top \) and \( \eta(x_i) = u_i \) for all \( i \in [1,n] \).

The Separability Problem. Given two classes of languages \( C \) and \( S \), the separability problem for \( C \) by the separator class \( S \) is defined as follows: Given two languages \( I \) and \( E \) from the class \( C \), does there exist a separator \( S \in S \) such that \( I \subseteq S \) and \( E \cap S = \emptyset \).
3 Regular Separability of String Constraints

Let $\Sigma$ be an alphabet and $k, n$ be two natural numbers. A set $R$ of $n$-tuples of words over $\Sigma$ is said to be regular (REG) if there is a sequence of finite-state automata $A_{(i,1)}, \ldots, A_{(i,n)}$ for every $i \in [1,k]$ such that $R=\bigcup_{i=1}^{k}[L(A_{(i,1)}) \times \cdots \times L(A_{(i,n)})]$. The REG separability problem for string constraints consists in checking for two given string constraints $\Psi$ and $\Psi'$ over the string variables $x_1, \ldots, x_n$ whether there is a regular set $R \subseteq (\Sigma^*)^n$ such that $L(\Psi) \subseteq R$ and $R \cap L(\Psi') = \emptyset$. The regular separability problem is undecidable in general. This can be seen as an immediate corollary of the fact that the satisfiability problem of string constraints is undecidable [36] even for a simple formula of the form $(x, x) \in R(T)$ where $T$ is a transducer and $x$ is a string variable. To see why, consider $\Psi$ to be $(x, x) \in R(T)$ and $\Psi'$ such that $L(\Psi') = \Sigma^*$. It is easy to see that $\Psi'$ and $\Psi$ are separable by a regular set if $\Psi$ is unsatisfiable. In the following, we show a stronger result, namely that this undecidability still holds even for REG separability between two SL formulas.

- **Theorem 2.** The REG separability problem is undecidable even for SL string constraints.

4 PTL-Separability of String Constraints

Given the undecidability of REG separability, we focus on the separability problem using piece-wise testable languages (PTL). We show that the problem is in general undecidable and then we show its decidability in the case of SL formulas. The undecidability proof is exactly the same as in the case of the REG separability (since $\Sigma^*$ is a PTL) while the decidability proof is done by reduction to its corresponding problem for the class of Ordered Multi Pushdown Automata (OMPA) [32] (which we show its decidability). In the rest of this section, we first recall the definition of PTL and extend it to $n$-tuples of words. Then, we define the class of OMPAs and show the decidability of its separability problem by PTL. Finally, we show the decidability of the separability problem for SL formulas by PTL.

**Piece-wise testable languages.** Let $\Sigma$ be an alphabet. A piece-language is a regular language of the form $\Sigma^* a_1 \Sigma^* a_2 \Sigma^* \cdots a_k \Sigma^*$ where $a_1, a_2, \ldots, a_k \in \Sigma$. The class of piecewise testable languages (PTL) is defined as a finite Boolean combination of piece languages [33]. We can define PTL for an $n$-tuple alphabet with $n \in \mathbb{N}$, as follows: The class of PTL over $n$-tuple words (denoted $n$-PTL) is defined as the finite Boolean combination of languages of the form $(\Sigma^*)^n v_1 (\Sigma^*)^n \cdots (\Sigma^*)^n v_k (\Sigma^*)^n$ where $v_i \in (\Sigma^*)^n$ for all $i \in [1, k]$.

**Ordered Multi Pushdown Automata.** Let $\Sigma$ be a finite alphabet and $n \geq 1$ a natural number. Ordered multi-pushdown automata extend the model of pushdown automata with multiple stacks. An $n$-Ordered Multi Pushdown Automaton (OMPA or n-OMPA) is a tuple $\mathcal{A} = (Q, \Sigma, \Gamma, \delta, Q_0, F)$ where (1) $Q, Q_0$ and $F$ are finite sets of states, initial states and final states, respectively, (2) $\Gamma$ is the stack alphabet and it contains the special symbol $\bot$, and (3) $\delta$ is the transition relation. OMPA are restricted in a sense that pop operations are only allowed from the first non-empty stack. A transition in $\delta$ is of the form $(q, \bot, \ldots, \bot, A_j \epsilon, \ldots, \epsilon) \rightarrow (q', \gamma_1, \ldots, \gamma_n)$ where $A_j \in \Gamma_*$ represents the symbol that will be popped from the stack $j$ on reading the input symbol $a \in \Sigma_*$, and $\gamma_i \in \Gamma^*$ represents the sequence of symbols which is going to be pushed on the stack $i$. The condition that $A_1 = \ldots = A_{j-1} = \bot$ (resp. $A_{j+1} = \ldots = A_n = \epsilon$) corresponds to the fact that the stacks $1, \ldots, j-1$ (resp. $j+1, \ldots, n$) are required to be empty (resp. inaccessible).

A configuration of $\mathcal{A}$ is of the form $(q, w, \alpha_1, \ldots, \alpha_n)$ where $q \in Q$, $w \in \Sigma^*$ and $\alpha_1, \ldots, \alpha_n \in (\Gamma \setminus \{\bot\})^* \cdot \{\bot\}$. The transition relation $\rightarrow$ between the set of configurations of $\mathcal{A}$ is defined as follows: Given two configurations $(q, w, \alpha_1, \ldots, \alpha_n)$ and
As a first step, we show that the class of OMPL forms a full trio [20]. This equivalence has been shown for full trios in [20] (see Lemma 3).

Proof. Lemma 4. ▶

Remark 7. The proof follows from Lemmas 5, 4 and [20], from where we know that the diagonal problem, which, given as input, a language asks whether each $n$-tuple $(m_1, ..., m_n) \in \mathbb{N}^n$ is dominated by another $n$-tuple $(d_1, ..., d_n)$ iff $m_i \leq d_i$ for all $1 \leq i \leq n$. The diagonal problem for $\mathcal{C}$ is the decision problem, which, given as input, a language $L$ from $\mathcal{C}$ asks whether each $n$-tuple $(m_1, ..., m_n) \in \mathbb{N}^n$ is dominated by some Parikh image of $L$.

Lemma 3. The class of OMPL forms a full trio.

To connect the PTL separability problem of SL string constraints to that of OMPL, we first use Lemma 2. Lemma 2 states that the PTL separability problem for OMPL is equivalent to the diagonal problem for OMPL. We recall the diagonal problem [20]. Fix a class of languages $\mathcal{C}$ as above and a language $L \in \mathcal{C}$ over alphabet $\Sigma = \{a_1, ..., a_n\}$. Assume an ordering $a_1 < \cdots < a_n$ on $\Sigma$. For $a \in \Sigma$ and $w \in L$, let $\#_a(w)$ denote the number of occurrences of $a$ in $w$. The Parikh image of $w$ is the $n$-tuple $(\#_{a_1}(w), ..., \#_{a_n}(w))$. The Parikh image of $L$ is the set of all Parikh images of words in $L$. An $n$-tuple $(m_1, ..., m_n) \in \mathbb{N}^n$ is dominated by another $n$-tuple $(d_1, ..., d_n)$ iff $m_i \leq d_i$ for all $1 \leq i \leq n$. The diagonal problem for $\mathcal{C}$ is the decision problem, which, given as input, a language $L$ from $\mathcal{C}$ asks whether each $n$-tuple $(m_1, ..., m_n) \in \mathbb{N}^n$ is dominated by some Parikh image of $L$.

Lemma 4. The PTL-separability and diagonal problems are equivalent for OMPL.

Proof. This equivalence has been shown for full trios in [20] (see Lemma 3).

Lemma 5. Each language $L$ in OMPL has a semilinear Parikh image.

Theorem 6. Given two OMPAs $A_1$ and $A_2$, checking whether there is a PTL $L$ such that $\mathcal{L}(A_1) \subseteq L$ and $L \cap \mathcal{L}(A_2) = \emptyset$ is decidable.

Proof. The proof follows from Lemmas 5, 4 and [20], from where we know that the diagonal problem is decidable for classes of languages having semilinear Parikh images.

Remark 7. For the case of 1–OMPA, the PTL separability problem is already known to be decidable [20] but its complexity is still an open problem.

4.1 From SL formula to OMPA

In the following, we show that the $n$-PTL separability problem for SL formulas can be reduced to the PTL separability problem for OMPLs. To that aim, we proceed as follows: First, we show how to encode an $n$-tuple of words $(\in (\Sigma^*)^n)$ as a word over $(\Sigma \cup \{\#\})^*$. Then, we show how to encode the set of solutions of an atomic relational constraint $(x, t) \in \mathcal{R}(T)$ using
the stacks of an OMPA. Finally, we construct an OMPA that accepts exactly the language of a given SL formula $\Psi$. This construction will make use of the constructed OMPAs that encode the set of atomic relational constraints appearing in $\Psi$. Let $\Sigma$ be an alphabet.

**Encoding an $n$-tuple of words.** Let $n$ be a natural number. We assume w.l.o.g. that the special symbol $\#$ does not belong to $\Sigma$. We define the function $\text{Encode}$ that maps any $n$-tuple word $w = (w_1, \ldots, w_n) \in (\Sigma^*)^n$ to the word $w_1 \# w_2 \# \cdots \# w_n$.

**From SL atomic relational constraints to OMPAs.** Let $x_1, x_2, \ldots, x_n$ be a sequence of string variables. Let $P_i$ be a relational constraint of the form $(x_i, t_i) \in \mathcal{R}(T_i)$ such that if a variable $x_i$ is appearing in the term $t_i$, then $j > i$. In the following, we show that we can construct an OMPA $A_i$ with $(3n + |t_i| + 2 - 3i)$ stacks such that if $A_i$ starts with a configuration where the first $(n - i)$ stacks contain, respectively, the evaluations $\eta(x_1), \ldots, \eta(x_n)$ (and all the other stacks are empty), then it can compute an evaluation $\eta(x_i)$ of the variable $x_i$ such that: (1) $(\eta(x_i), \eta(t_i)) \in \mathcal{R}(T_i)$ and the evaluations $\eta(x_1), \ldots, \eta(x_n)$ are stored in the last $n - i + 1$ stacks of $A_i$. Such an OMPA $A_i$ will be used as a gadget when constructing the OMPA $A$ that accepts exactly the language of a given SL formula $\Psi$.

**Lemma 8.** We can construct an OMPA $A_i = (Q_i, \Sigma, \{\bot\} \cup \Sigma, \delta_i, \{q_i^{\text{init}}\}, \{q_i^{\text{final}}\})$ with $(3n + |t_i| + 2 - 3i)$-stacks such that for every $u_1, \ldots, u_n \in \Sigma^*$, we have $(q_i^{\text{init}}, \epsilon, u_1\perp, \ldots, u_n\perp, \bot, \ldots, \bot) \rightarrow^* (q_i^{\text{final}}, \epsilon, \bot, \ldots, \bot, u_i\perp, u_{i+1}\perp, \ldots, u_n\perp)$ iff $(\eta(x_i), \eta(t_i)) \in \mathcal{R}(T_i)$ with $\eta(x_j) = u_j$ for all $j \in [i, n]$.

**Proof.** In the proof, we omit the input $\epsilon$ from the OMPA configurations, and only write the state, and stack contents. Let us assume that the string term $t_i$ is of the form $y_1y_2 \cdots y_{|t_i|}$. Observe that $y_j \in \{x_1, \ldots, x_n\}$. The OMPA $A_i$ proceeds in phases starting from the configuration $(q_i^{\text{init}}, u_{i+1}\perp, \ldots, u_n\perp, \bot, \ldots, \bot)$. To begin, stacks $1$ to $n - i$ contain $u_1, \ldots, u_n$, the evaluations of $x_{i+1}, \ldots, x_n$, and all other stacks are empty. The computation proceeds in 4 phases. The stacks indexed $1, \ldots, n - i$ and $n - i + 1, \ldots, 2n - 2i$ will be used in the first phase below. The second phase uses stacks indexed $n - i + 1, \ldots, 2n - 2i$ and $2n - 2i + 1, \ldots, 2n - 2i + |t_i|$ along with the last $n - i$ stacks indexed $2n - 2i + |t_i| + 3$ to $3n - 3i + |t_i| + 2$. In the third phase, stacks indexed $2n - 2i + 1, \ldots, 2n - 2i + |t_i|, 2n - 2i + |t_i| + 1$ are used. In the last phase, stacks indexed $2n - 2i + |t_i| + 1$ and $2n - 2i + |t_i| + 2$ are used. At the end of the 4 phases, stacks indexed $2n - 2i + |t_i| + 2, \ldots, 3n - 3i + |t_i| + 2$ hold the evaluations of $x_1, x_{i+1}, \ldots, x_{n}$, and all other stacks are empty.

**Phase 1.** The OMPA $A_i$ pops the symbols, one by one, from the first $(n - i)$-stacks $1, \ldots, n - i$ and pushes them into the stacks from index $(n - i + 1)$ to $(2n - 2i)$, respectively. At the end of this phase, the new configuration of the OMPA $A_i$ is $(q_i^{\text{init}}, \bot, \ldots, \bot, u_{i+1}\perp, \ldots, u_n\perp, \bot, \ldots, \bot)$. That is, stacks $n - i + 1, \ldots, 2n - 2i$ have $u_{i+1}, \ldots, u_n$, while all other stacks are empty.

**Phase 2.** We do two things. (1) the contents of the $n - i$ stacks $n - i + 1, \ldots, 2n - 2i$ are moved (in reverse) into the $n - i$ stacks $2n - 2i + |t_i| + 3, \ldots, 3n - 3i + |t_i| + 2$. This results in the stacks $2n - 2i + |t_i| + 3, \ldots, 3n - 3i + |t_i| + 2$ containing $u_{i+1}, \ldots, u_n$. (2) If $y_j$ appearing in $t_i$ is the variable $x_{i+j}$, then the content of stack $n - i + \ell$ (with $n - i + 1 \leq n - i + \ell \leq 2n - 2i$) is also moved (in reverse) to stack $2n - 2i + j$, $1 \leq j \leq |t_i|$. This results in stack $2n - 2i + j$...
containing \(u_{i+\ell}\). Thus, at the end of (1), (2), the stacks \(n-i+1, \ldots, 2n-2i\) are empty, the stack \(2n-2i+|t_1|+\ell+2\) contains \(u_{i+\ell}\), the evaluation of \(x_{i+\ell}\) for \(\ell \geq 1\), while stack \(2n-2i+k\) for \(1 \leq k \leq |t_1|\) contains \(u_{i+m}\) if \(yk = x_{i+m}\). The two stacks \(2n-2i+|t_1|+1\) and \(2n-2i+|t_1|+2\) are empty at the end of this phase. Stack contents of \(2n-2i+k, 1 \leq k \leq |t_1|\) are referred to as \(v_k\) in the figure.

**Phase 3.** The OMPA \(A_i\) mimics the transducer \(T_i\). The current state of \(A_i\) is the same as the current state of the simulated transducer. Each transition of \(T_i\) of the form \((q, (a, b), q')\) is simulated by (1) moving the state of \(A_i\) from \(q\) to \(q'\), (2) pushing the symbol \(a\) into the stack \((2n-2i+|t_1|+1)\), and (3) popping the symbol \(b\) from the first non-empty stack having an index between \(2n-2i+1\) to \(2n-2i+|t_1|\). Recall that the stacks \(2n-2i+1\) to \(2n-2i+|t_1|\) contain the evaluations of \(y_1, \ldots, y_{|t_1|}\), for \((\eta(x_i), \eta(y_1), \eta(y_2), \ldots, \eta(y_{|t_1|})) \in R(T_i)\).

When the current state of \(A_i\) is in a final state of \(T_i\) and the stacks from index \(2n-2i+1\) to \(2n-2i+|t_1|\) are empty, then we know that \(\eta(y_1) \ldots \eta(y_{|t_1|})\) is indeed related by \(T_i\) on \(\eta(x_i)\). Then \(A_i\) changes its state to \(q_{i}^{\text{final}}\).

**From SL formula to OMPAs.** In the following, we first construct an OMPA that accepts the encoding of the set of solutions of an SL formula.

**Lemma 9.** Given an SL formula \(\Psi\), with \(x_1, \ldots, x_n\) as its set of variables, it is possible to construct an OMPA \(A\) such that \(L(A) = \text{Encode}(L(\Psi))\).

**Proof.** Let us assume that \(\Psi\) is of the form \(\bigwedge_{i=1}^{n} x_i \in L(A_i) \wedge \bigwedge_{i=1}^{k} \varphi_i\), where \(\varphi_1, \ldots, \varphi_k\) are relational constraints such that \(\varphi_i\) is of the form \((x_i, t_i) \in R(T_i)\). The OMPA \(A\) will have \((n-k + \sum_{i=1}^{k} (2n-2i+2+|t_i|))\) stacks. \(A\) first guesses an evaluation for the variables \(x_{k+1}, \ldots, x_n\) in the first \(n-k\) stacks and then starts simulating the OMPA \(A_k\) (see Lemma 8 for the definition of \(A_k\)) in order to compute a possible evaluation of the variable \(x_1\) such that the relational constraint \((x_k, t_k) \in R(T_k)\) holds for that evaluation. After this step, the stacks from index \((2n-2k+|t_k|+2)\) to \((3n-3k+|t_k|+2)\) contain the evaluation of the string variables \(x_{k+1}, \ldots, x_n\), and all remaining stacks are empty. Now \(A\) can start the simulation of the OMPA \(A_{k-1}\) (Lemma 8) in order to compute a possible evaluation of the variable \(x_{k-1}\) such that \((x_{k-1}, t_{k-1}) \in R(T_{k-1})\) holds for that evaluation. At the start of the simulation of \(A_{k-1}\) by \(A\), the \(n-k+1\) stacks (indexed \((2n-2k+|t_k|+2)\) to \((3n-3k+|t_k|+2)\)) contain the evaluations of \(x_{k+1}, \ldots, x_n\), and the next \(2n-2(k-1)+|t_{k-1}|+2\)
stacks are used to simulate phases 2-4 of \( A_{k-1} \). At the end of this, the \( n-k+2 \) stacks backwards from the stack indexed \((3n-3k+|t_k|+2)+2n-2(k-1)+|t_{k-1}|+2\) contain the evaluations of \( x_{k-1}, \ldots, x_n \). Now, \( A \) simulates \( A_{k-2}, \ldots, A_n \) in the same way. At the end of this simulation phase, the last \( n \)-stacks of \( A \) contain an evaluation of the string variables \( x_1, \ldots, x_n \) that satisfies \( \bigwedge_{i=1}^{k} \varphi_i \). Let us assume that the current configuration of \( A \) at the end of this is of the form \((q^{\text{final}}, \bot, \ldots, \bot, \underline{u_1}, \underline{u_2}, \ldots, \underline{u_n})\). Then, \( A \) starts popping, one-by-one, from the \( n \)-th stack from the last and outputs the read stack symbol \( \in \Sigma \) while ensuring that the evaluation \( u_1 \) of \( x_1 \) belongs to \( \mathcal{L}(A_1) \). When the \( n \)-th stack from the last is empty, \( A \) outputs the special symbol \( \# \). Then, \( A \) does the same for the \( i \)-th stack from last, with \( i \in [1, n-1] \), which contains the evaluation of \( x_{i+1} \). If \( A \) succeeds to empty all stacks, then this means that the evaluation \( \eta \) which associates to the variable \( x_i \) the word \( u_i \) for all \( i \in [1, n] \) satisfies \( \bigwedge_{i=1}^{n} x_i \in \mathcal{L}(A_i) \). Hence, \( u_1 \#u_2\# \cdots \#u_n \) is accepted by \( A \) iff \( \eta \models \Psi \).

The following lemma shows that the PTL-separability problem for SL formulas can be reduced to the PTL-separability problem for OMPLs.

\textbf{Lemma 10.} Let \( \Psi_1 \) and \( \Psi_2 \) be two SL formulae with \( x_1, \ldots, x_n \) as their set of variables. Let \( A_1 \) and \( A_2 \) be two OMPAs such that \( \mathcal{L}(A_1) = \text{Encode}(\mathcal{L}(\Psi_1)) \) and \( \mathcal{L}(A_2) = \text{Encode}(\mathcal{L}(\Psi_2)) \). \( \Psi_1, \Psi_2 \) are \( n \)-PTL separable iff \( A_1, A_2 \) are PTL separable.

As an immediate corollary of Theorem \( 6 \) Lemma \( 10 \) we obtain our main result:

\textbf{Theorem 11.} The \( n \)-PTL separability problem of SL formulae is decidable.

5 \textbf{PosPTL-Separability of String Constraints}

In this section, we address the separability problem for string constraints by a sub-class of PTL, called positive piece-wise testable languages (PosPTL). A language is in PosPTL iff it is defined as a finite positive Boolean combination (i.e., union and intersection but no complementation) of piece-languages. Given a natural number \( n \in \mathbb{N} \), this definition can naturally be extended to \( n \)-tuples of words in the straightforward manner (as in the case of PTL) to obtain the class of \( n \)-PosPTL. In the following, we first provide a necessary and sufficient condition for the \( n \)-PosPTL separability problem of any two languages.

\textbf{Theorem 12.} Two languages \( I \) and \( E \) are \( n \)-PosPTL separable iff \( I \uparrow \cap E = \emptyset \) iff \( I \cap E \downarrow = \emptyset \).

The rest of this section is structured as follows: First, we show that the PosPTL separability is decidable for OMPLs; in the particular case of CFLs, this problem is PSPACE-complete. Then, we use the encoding of SL formulas to OMPAs (as defined in Section \( 4 \)) and show that \( n \)-PosPTL separability of SL formulas reduces to the PosPTL-separability of corresponding OMPLs. Finally, we consider the PosPTL-separability problem for a subclass of SL formulas, called right sided SL formulas. We show that the PosPTL separability problem for this subclass is PSPACE-HARD and is in EXPSPACE.

5.1 \textbf{PosPTL separability of SL formulas}

First, we show that the PosPTL separability for OMPLs is decidable.

\textbf{Theorem 13.} PosPTL separability of OMPLs is decidable.
Proof. Consider $\mathcal{C}$ to be the class of OMPLs in Theorem 12. Let $I$ and $E$ be two languages belonging to $\mathcal{C}$ as stated in Theorem 12. Then, the set $\operatorname{min}(I\uparrow)$ is effectively computable as an immediate consequence of the Generalized Valk-Jantzen construction [6]. The main idea behind this construction is to start with an empty minor set $M$ (so to begin, $M \subseteq I$) and keep adding new words $w \in I$ to $M$ if $w$ is not already in $M\uparrow$. Before adding a new word, we need to test that $I \cap M\uparrow \neq \emptyset$ (the complement of $M\uparrow$ intersects with $I$). This test is decidable since (i) OMPLs are closed under intersection with regular languages and (ii) the emptiness problem for OMPA is decidable. At each step, we remove all the non-minimal words from $M$ (since $M$ is finite). The algorithm terminates due to the Higman’s Lemma 26 (the minor of an upward closed set is finite). When the algorithm terminates, $I \subseteq M\uparrow$ and thus $I\uparrow \subseteq M\uparrow$. By construction, $M \subseteq I$ and $M\uparrow \subseteq I\uparrow$. Thus, $M\uparrow = I\uparrow$. Since $M$ is a minor set, we have $\operatorname{min}(I\uparrow) = M$. Using (i) and (ii), we obtain the decidability of checking the emptiness of $I\uparrow \cap E$, and thus PosPTL separability of OMPL is decidable. \hfill \qed

As mentioned in section 4, the complexity of PTL-separability for 1-OMPL is open; however, we show that the PosPTL separability problem for 1-OMPL is PSPACE-COMPLETE.

\textbf{Theorem 14.} The PosPTL-separability for CFLs is PSPACE-COMPLETE.

For the decidability of the $n$-PosPTL separability of SL formulas, we use the encoding of SL formulas to OMPAs (as defined in section 4), and show that the $n$-PosPTL separability of SL formulas reduces to the PosPTL separability of their corresponding OMPLs. The decidability of the $n$-PosPTL separability of SL formulas follows from Theorem 13.

\textbf{Lemma 15.} Given two SL formulas $\Psi$ and $\Psi'$, with $x_1, \ldots, x_n$ as their set of variables. Let $A$ and $A'$ be two OMPAs such that $L(A) = \operatorname{Encode}(L(\Psi))$ and $L(A') = \operatorname{Encode}(L(\Psi'))$. Then, $\Psi$ and $\Psi'$ are separable by an $n$-PosPTL iff $A$ and $A'$ are separable by a PosPTL.

As an immediate corollary of Lemma 13 and 15, we obtain the following theorem:

\textbf{Theorem 16.} The $n$-PosPTL separability problem of SL formulae is decidable.

5.2 PosPTL separability of right-sided SL formula

Unfortunately, the proof of Theorem 16 does not allow us to extract any complexity result. Therefore, we consider in this subsection a useful fragment of SL formulas, called right-sided SL formulas. Roughly speaking, an SL formula $\Psi$ is right-sided iff any variable appearing on the right-side of a relational constraint cannot appear on the left-side of any relational constraint. Let us formalize the notion of right-sided SL formulas. Let us assume an SL formula $\Psi$ of the form $\bigwedge_{i=1}^{n} x_i \in \mathcal{L}(A_i) \land \bigwedge_{i=1}^{k} (x_i, t_i) \in \mathcal{R}(T_i)$ with $x_1, \ldots, x_n$ as set of variables. Then, $\Psi$ is said to be right-sided if none of the variables $x_1, \ldots, x_k$ appear in any of $t_1, \ldots, t_k$. We call $x_{k+1}, \ldots, x_n$ (resp. $x_1, \ldots, x_k$) independent (resp. dependent) variables. Observe that the class of SL formulas with functional transducers can be rewritten as right-sided SL formulas. A transducer $T$ is functional if for every word $w$, there is at most one word $w'$ such that $(w', w) \in \mathcal{R}(T)$ ($T$ computes a function). An example of a functional transducer is the one implementing the identity relational constraint (allowing to express the equality $x = t$).

In the following, we show that the PosPTL-separability problem for right-sided SL formulas is in EXPSPACE. To show this result, we will reduce the PosPTL-separability problem for right-sided SL formulas to its corresponding problem for two-way transducers.

Two way transducers. Let $\Sigma$ be a finite input alphabet and $\triangleright, \triangleleft$ be two special symbols not in $\Sigma$. We assume that every input string $w \in \Sigma^*$ is presented as $\triangleright w \triangleleft$, where $\triangleright, \triangleleft$ serve
as left and right delimiters that appear nowhere else in \( w \). We write \( \Sigma_{i-1} = \Sigma \cup \{ \sqcup, \sqcap \} \). A two-way automaton \( \mathcal{A} = (Q, \Sigma, \delta, I, F) \) has a finite set of states \( Q \), subsets \( I, F \subseteq Q \) of initial and final states and a transition relation \( \delta \subseteq Q \times \Sigma_{i-1} \times Q \times \{-1, 1\} \). The -1 represents the reading head moves to left after taking the transition while a 1 represents that it moves to right.

The reading head cannot move left if it is on \( \sqcap \), and cannot move right when it is on \( \sqcup \). A configuration of \( \mathcal{A} \) on reading \( w' = \sqcup w \sqcup \) is represented by \( (q, i) \) where \( q \in Q \) and \( i \) is a position in the input, \( 1 \leq i \leq |w| + 2 \), which will be read in state \( q \). An initial configuration is of the form \( (q_0, \sqcup) \) with \( q_0 \in I \) and the reading head on \( \sqcup \). If \( w' = w_1aw_2 \) and the current configuration is \( (q, |w_1| + 1) \), and \( (q, a, q', -1) \in \delta \), then there is a transition from the configuration \( (q, |w_1| + 1) \) to \( (q', |w_1|) \) (hence \( a \neq \sqcup \)). Likewise, if \( (q, a, q', 1) \in \delta \), we obtain a transition from \( (q, |w_1| + 1) \) to \( (q', |w_1| + 2) \). A run of \( \mathcal{A} \) on reading \( \sqcup w \sqcup \) is a sequence of transitions; it is accepting if it starts in an initial configuration and ends in a configuration of the form \( (q, |w| + 2) \) with \( q \in F \) and the reading head on \( \sqcup \). The language of \( \mathcal{A} \) (denoted \( L(\mathcal{A}) \)) is the set of all words \( w \in \Sigma^* \) s.t. \( \mathcal{A} \) has an accepting run on \( \sqcup w \sqcup \).

We extend the definition of a two-way automaton \( \mathcal{A} = (Q, \Sigma, \delta, I, F) \) into a two-way transducer (2NFT) \( \mathcal{A} = (Q, \Sigma, \Gamma, \delta, I, F) \) where \( \Gamma \) is a finite output alphabet. The transition relation is defined as a finite subset \( \delta \subseteq Q \times \Sigma_{i-1} \times \Gamma^* \times \{-1, 1\} \). The output produced on each transition is appended to the right of the output produced so far. \( \mathcal{A} \) defines a relation \( \mathcal{R}(\mathcal{A}) = \{ (w, u) \mid w \text{ is the output produced on an accepting run of } u \} \). The acceptance condition is the same as in two-way automata. Sometimes, we use the macro-notation \((p, q, a, s, \alpha, 0)\) to denote a sequence of consecutive transitions \((p, a, s, \alpha, d' \) \) and \((s, b, q, \epsilon, d')\) in \( \delta \) with \( d + d' = 0 \), \( b \in \Sigma_{i-1} \) and \( s \) is an extra intermediary state of \( \mathcal{A} \) that is not used anywhere else (and that we omit from the set of states of \( \mathcal{A} \)).

**PosPTL separability of 2NFT.** In the following, we study the PosPTL separability of 2NFT. We define here the notion of visiting sequences (similar to crossing sequences of 2NFT [S]), which will be used in the proof of Lemma 17. Let \( w = \sqcup a_1 \ldots a_n \sqcup \) be an input word and let \( \rho \) be a run of the 2NFT on \( w \). A visiting sequence at a position \( x \) in a run \( \rho \) of \( w \) captures the states visited in order in the run, each time the reading head is on position \( x \), along with the information pertaining to the direction of the outgoing transition from that state.

For example, in run \( \rho \), if position \( x \) is visited for the first time in state \( q \), and the outgoing transition chosen in \( \rho \) from \( q \) during that visit had direction \( +1 \), then \( q^+ \) will be the first entry in the visiting sequence. For a run \( \rho \), the visiting sequence at a position \( x \) is defined as the tuple \( \rho|_x = (q^1_1, q^2_1, \ldots, q^h_1) \) of states that have, in order, visited position \( x \) in \( \rho \), and whose outgoing transitions had direction \( d_1, \ldots, d_h \). In the example, the visiting sequence at position 2 is \( (q^1_1, q^-_1, q^+_1) \), while those at 1 and 3 respectively are \( (q^-_1, q^+_1) \) and \( (q^+_2, q^-_2) \).

**Lemma 17.** Given a 2NFT \( \mathcal{T} \), if \((v, u)\) is in \( \min(\mathcal{R}(\mathcal{T})) \) then \(|u| \) and \(|v| \) are of at most exponential length in the size of \( \mathcal{T} \).

**Proof.** In the following, we show that, if \((v, u) \in \min(\mathcal{R}(\mathcal{T})) \), then \(|u| \leq \text{in}_{\max} = \sum_{i=1}^{|Q|} (|2(Q)|)^i \cdot |\Sigma| \) and \(|v| \leq \text{out}_{\max} = \sum_{i=1}^{|Q|} (|2(Q)|^i \cdot |\Sigma| \cdot |Q| \cdot \gamma_{\max}) \) where \( \gamma_{\max} \) represents the maximum length of an output on any transition in \( \mathcal{T} \). To show this result, we need to define normalized runs as follows: A run is normalized if it visits each state at most once on each position \( x \). In the following, we show that \((v, u)\) can be generated by a normalized run \( \rho \). Assume that \((v, u)\) is accepted by a run \( \rho' \) which is not normalized. Then, we will have, in \( \rho' \), two visits to some position \( x \) of the word in the same state \( q \). After the first visit to position \( x \) in state \( q \), the transducer has explored some positions till its
On the Separability Problem of String Constraints

The length of visiting sequences in a normalized run is \(|Q|\) and hence the number of visiting sequences is at most exponential in \(|Q|\), precisely it is \(\leq \sum_{i=1}^{|Q|} (2|Q|)^i\). Suppose \(|u| > \text{in}_{\text{max}}.\) Then there exists a visiting sequence which is repeated on reading the same input symbol in the accepting run of \(u\), at positions \(i \neq j\). By deleting the part between the \(i\)th and \((j - 1)\)th position, we again obtain an accepting run over a word \(u'\), which is a strict subword of \(u\), and whose output \(v'\) is also a subword (may not be strict) of \(v\), a contradiction to \((v, u) \in \min(\mathcal{R}(T))\). Now suppose \(|u| \leq \text{in}_{\text{max}}\) on reading each position of \(u\), at most \((|Q|\text{in}_{\text{max}})\) symbols can be produced. Hence, we have \(|v| \leq (|Q| \cdot \text{out}_{\text{max}} \cdot |u|)\). ▶

From Theorem 12 and Lemma 17 the following result holds:

**Lemma 18.** The 2-PosPTL separability problem for 2NFT is in \textsc{ExpSpace}.

**Proof.** Using Theorem 12 we know that \(\mathcal{R}(T_1) \cap \mathcal{R}(T_2) = \emptyset\) iff \(T_1\) and \(T_2\) are 2-PosPTL separable. Here is an \textsc{NExpSpace} algorithm.

1. Guess some \((v, u)\) s.t. the lengths of \(v, u\) are at most as given by the proof of Lemma 17.
2. Check if \((v, u) \in \mathcal{R}(T_1)\). If yes, then do (3). Else exit.
3. Check if \((v, u) \uparrow \cap \mathcal{R}(T_2) \neq \emptyset\). The guessed word \((v, u)\) has exponential length in the size of \(T_1\). To check if \((v, u) \in \mathcal{R}(T_1)\), we construct another transducer \(T'_1\) that first checks that its input word is \(u\), then it comes back to \(\vdash\) and starts simulating \(T_1\), while also keeping track, longer and longer prefixes of \(v\). We then compare those prefixes with the output produced by \(T_1\). This gives rise to exponentially many states (maintaining prefixes of \(u\) and \(v\)) and we finish when \(T_1\) enters an accepting state, and at the same time, the produced word is \(v\). Since \(\mathcal{R}(T'_1) = \{(v, u) \in \mathcal{R}(T_1)\}\) by construction, checking if \((v, u) \in \mathcal{R}(T_1)\) can be reduced to the emptiness problem of \(T'_1\). After this, we check the emptiness of \((v, u) \uparrow \cap \mathcal{R}(T_2)\). This is done as follows. First, construct automata \(A_u, A_v\) accepting languages \{\(u\)\} and \{\(v\)\} respectively. The number of states of \(A_u, A_v\) are exponential in the number of states of \(T_1\), since the lengths of \(u, v\) have this bound. Then, we construct a transducer \(T'_2\) such that \(\mathcal{R}(T'_2) = \{(v, u) \uparrow \cap \mathcal{R}(T_2)\}\) in a similar manner as \(T'_1\). \(T'_2\) reads the input word while simulating \(A_v\). On entering an accepting state of \(A_v\), it comes back to \(\vdash\). Then it simulates \(T_2\), and, on the outputs produced, simulates \(A_u\). If \(A_u\) enters an accepting state at the same time \(T_2\) accepts, then we are done. The state space of \(T'_2\) is exponential in the states of \(T_1\) and linear in the states of \(T_2\). Since \(\mathcal{R}(T'_2) = \{(v, u) \uparrow \cap \mathcal{R}(T_2)\}\), checking the emptiness of \((v, u) \uparrow \cap \mathcal{R}(T_2)\) can be reduced to checking the emptiness problem of \(T'_2\). The emptiness problem for 2NFT is known to be \textsc{PSPACE}-COMPLETE [40]. Thus, in our case, the emptiness of \(T'_1\) and \(T'_2\) can be achieved in space exponential in \(T_1\). Since we can handle the second and third steps in exponential space, we obtain an \textsc{NExpSpace} algorithm. By Savitch’s Theorem, we obtain the \textsc{ExpSpace} complexity. ▶

**From Right-sided SL formulas to 2NFT.** Hereafter, we show how to encode the set of solutions of a right-sided SL formula using 2NFT. Let \(\Sigma\) be an alphabet and \(\# \notin \Sigma\).

**Lemma 19.** Let \(\Psi\) be a right-sided SL formula over \(\Sigma\), with \(x_1, x_2, \ldots, x_n\) as its set of variables. Then, it is possible to construct, in polynomial time, a 2NFT \(A_\Psi\) such that
\( \mathcal{R}(A_\Phi) = \{(u_1 \# u_2 \# \cdots \# u_n, w_1 \# u_2 \# \cdots \# u_n) \mid \text{Encode}(\mathcal{L}(\Psi)) \text{ and } w_1 = u_i \text{ if } x_i \text{ is an independent variable} \} \)

**Proof.** Let us assume that \( \Psi \) is of the form \( \bigwedge_{i=1}^{n} y_i \in \mathcal{L}(A_i) \land \bigwedge_{i=1}^{k} (y_i, t_i) \in \mathcal{R}(T_i) \) with \( y_1, \ldots, y_n \) is a permutation of \( x_1, \ldots, x_n \). Let \( \pi : [1, n] \to [1, n] \) be the mapping that associates to each index \( i \in [1, n] \), the index \( j \in [1, n] \) s.t. \( x_i = y_j \) (or \( x_i = y_{\pi(i)} \)). We construct \( A_\Psi \) as follows: \( A_\Phi \) reads \( n \) words over \( \Sigma \) separated by \( \# \) as input. We explain hereafter the working of \( A_\Phi \) when it produces the assignment for \( x_1 \) (the other variables are handled in similar manner).

- Assume that \( x_1 \) is a dependent variable. Let \( \varphi_{(1)}(1) \in \mathcal{R}(T_{(1)}) \), with \( t_{(1)} = x_{i_1} x_{i_2} \ldots x_{i_k} \) and \( x_{i_j} \in \{y_{k+1}, y_{k+2}, \ldots, y_n\} \) for all \( j \). First, \( A_\Phi \) reads \( x_{i_1} \) i.e. the first variable in \( t_{(1)} \). To read \( x_{i_1} \), it skips \( (i_1 - 1) \) many blocks separated by \#s of the input, and comes to \( w_{i_1} \). On the first symbol of \( w_{i_1} \), \( A_\Phi \) starts mimicking transitions of \( T_{(1)} \) from its initial state, while producing the same output as \( T_{(1)} \). On the same output, \( A_\Phi \) mimics the transitions of \( A_{\pi(1)} \) starting from the initial state to check the membership constraint of \( y_{\pi(1)} \). This can be done by a product construction between \( A_{\pi(1)} \) and \( T_{(1)} \). For instance, \( A \) will have a transition \( ((p, q), a, (p', q'), b, 1) \) (resp. \( ((p, q), a, (p', q'), b, 0) \)), if there are transitions \( (p, (b, a), p') \) (resp. \( (p, (b, e), p') \)) in \( T_{(1)} \) and \( (q, b, q') \) in \( A_{\pi(1)} \). If it reaches \# or \( \vdash \) in the input, it remembers the current states of \( T_{(1)} \) and \( A_{\pi(1)} \), say \( (p_1, q_1) \) in its control state. Next, \( A_\Phi \) reads \( x_{i_2} \) in the input. To read \( x_{i_2} \), \( A_\Phi \) moves to \( \vdash \) and then changes direction. As before it reaches \( x_{i_2} \) by skipping \( (i_2 - 1) \) many \#s, and starts reading the input (the first symbol of \( w_{i_2} \)) from the state \( (p_1, q_1) \) stored in the finite control. Transitions are similar to explained above. This procedure is repeated to read \( x_{i_3} \ldots x_{i_n} \). After reading \( x_{i_n} \), if the next state contains the pair \( (p_c, q_c) \), where \( p_c \) (resp. \( q_c \)) is a final state of \( T_{(1)} \) (resp. \( A_{\pi(1)} \)), we can say that the output produced till now satisfies \( \varphi_{(1)}(1) \) and \( y_{\pi(1)} \in \mathcal{L}(A_{\pi(1)}) \).

- Assume now that \( x_1 \) is an independent variable, then \( A_\Phi \) needs to read \( x_1 \). We need a single pass of the input which verifies if the first block corresponding to value of \( x_1 \) in input indeed satisfies its corresponding membership constraint. During this pass \( A_\Phi \) mimics transitions of \( A_{\pi(1)} \) starting from its initial states, and outputs the same letter as input.

The above procedure is repeated for all variables from \( x_2 \) to \( x_n \). After each pass, \( A \) moves to \( \vdash \) and then changes direction. Irrespective of whether \( x_i \) is dependent or not, while going from \( x_i \) to \( x_{i+1}, i \in [1, n - 1] \), \( A \) outputs a \# as separator. From the description above, it can be seen that if \( x_i \) is independent, then its evaluation \( u_i \) given as the \( i \)th block of the input is equal to the output \( w_i \), and if \( x_i \) is a dependent variable, then the output block \( w_i \) is the output of \( T_{(i)} \).

Notice that the above construction of 2NFT relies on the right-sidedness: if a variable \( x_i \) appears in the output of \( T_i \) and also in the input of \( T_k \) for some \( k \), then we will have to store the produced evaluation of \( x_i \) in order to use it later on when processing \( T_k \). However, there is no way to store the produced evaluation of \( x_i \) or compare it with its input evaluation. Next, we show that the PosPTL separability problem for right-sided formulas can be reduced to its corresponding problem for 2NFT.

| Lemma 20. Let \( \Psi_1 \) and \( \Psi_2 \) be two right-sided SL formula, with \( x_1, \ldots, x_n \) as their set of variables. Let \( A_{\Psi_1} \) and \( A_{\Psi_2} \) be the two 2NFT's encoding, respectively, the set of solutions of \( \Psi_1 \) and \( \Psi_2 \) (as described in Lemma 17). Then, the two formulae \( \Psi_1 \) and \( \Psi_2 \) are separable by \( n \)-PosPTL iff \( \mathcal{R}(A_1) \) and \( \mathcal{R}(A_2) \) are separable by a 2-PosPTL.

Proof. Let \( \mathcal{R}(A_{\Psi_1}) \) and \( \mathcal{R}(A_{\Psi_2}) \) be separable by a 2-PosPTL \( L \). By definition, \( L \) is a Boolean combination (except complementation) of piece languages of words over the two
On the Separability Problem of String Constraints

tuple alphabet $(\Sigma \cup \{\#\})^2$. We can assume w.l.o.g. that $L$ is the union of piece languages. This is possible since the intersection of two piece languages can be rewritten as a union of piece languages. Consider $L' = L \cap (R \times R)$, where $R$ is a regular language consisting of words having exactly $(n - 1)$ #s. We claim that $L'$ can be rewritten as the union of languages of the form $[L_1 \# L_2 \# \ldots \# L_n] \times [R_1 \# R_2 \# \ldots \# R_n]$ where the $L_i$s and $R_i$s are piece languages over $\Sigma$, and that $L'$ is also a separator of $R(A_{\Psi_1})$ and $R(A_{\Psi_2})$.

We prove this claim inductively. As a base case consider $L$ to be a piece language $((\Sigma \cup \{\#\}))^*(a_1, b_1)((\Sigma \cup \{\#\}))^* \ldots ((a_m, b_m)((\Sigma \cup \{\#\}))^*)^2$. Let $S$ be a finite set containing only the minimal words of the form $(w, w')$ such that $a_1 a_2 \ldots a_n \preceq w$, $b_1 b_2 \ldots b_m \preceq w'$, and the symbol # appears exactly $(n - 1)$-times in $w$ and $w'$. Thus $L \cap (R \times R) = \bigcup_{(a'_1, \ldots, a'_n, b'_1, \ldots, b'_m) \in S} [\Sigma^* a'_1 \Sigma^* a'_2 \ldots a'_n \Sigma^*] \times [\Sigma^* b'_1 \Sigma^* b'_2 \ldots b'_m \Sigma^*]$. So $L \cap (R \times R)$ is the union of piece languages of the form $[L_1 \# L_2 \# \ldots \# L_n] \times [R_1 \# R_2 \# \ldots \# R_n]$ where the $L_i$s and $R_i$s are piece languages over $\Sigma$. Now assume that $L$ is of the form $L_1 \cup L_2$. It is easy to see that $L \cap (R \times R)$ is equivalent to $(L_1 \cap (R \times R)) \cup (L_2 \cap (R \times R))$. Thus we can use our induction hypothesis to show that $L \cap (R \times R)$ is the union of languages of the form $[L_1 \# L_2 \# \ldots \# L_n] \times [R_1 \# R_2 \# \ldots \# R_n]$ where $L_i$s and $R_i$s are piece languages over $\Sigma$. Next we prove that $L'$ is a separator of $R(A_{\Psi_1})$ and $R(A_{\Psi_2})$. Indeed if $(v, u) \in R(A_{\Psi_1})$ and $(v, u) \in R(A_{\Psi_2})$. Since $L$ is a separator, we have $(v, u) \in L'$ and hence $(v, u) \in L'$. Suppose $(v, u) \in R(A_{\Psi_1}) \cap L'$, then $(v, u) \in L \cap R(A_{\Psi_1})$ since $(v, u) \in L'$, and $L' \subseteq L$, which is a contradiction with the assumption that $L$ is a separator.

Now we are in a condition to provide $n$-PosPTL separator for $L(\Psi_1)$ and $L(\Psi_2)$, using $L'$. Given a language of the form $[L_1 \# L_2 \# \ldots \# L_n] \times [R_1 \# R_2 \# \ldots \# R_n]$ where $L_i$s and $R_i$s are piece languages, we associate to it an $n$-PosPTL equivalent to $((L_1 \cap R_1) \times (L_2 \cap R_2) \times \ldots \times (L_n \cap R_n))$, i.e. the idea is to generate the $n$ dimensions in the $n$-PosPTL from the $n$ #-separated blocks in two dimensions. This definition is extended in the straightforward manner to union of piece languages. Let $K$ be the $n$-PosPTL associated to $L'$. $K$ is indeed a separator of $L(\Psi_1)$ and $L(\Psi_2)$: Suppose $v = (w_1, \ldots, w_n) \in L(\Psi_1)$, then $(w_1 \# \ldots \# w_n, w_1 \# \ldots \# w_n) \in R(A_{\Psi_1})$ (from the definition of $A_{\Psi_1}$). Since $L'$ is a separator, $(w_1 \# \ldots \# w_n, w_1 \# \ldots \# w_n) \in L'$. By construction of $K$, $(w_1, \ldots, w_n) \in K$. Assume $v = (w_1, \ldots, w_n) \in L(\Psi_2) \cap K$, then $(w_1 \# \ldots \# w_n, w_1 \# \ldots \# w_n) \in L'$. Since $L' \cap R(A_{\Psi_2}) = \emptyset$, then $(w_1 \# \ldots \# w_n, w_1 \# \ldots \# w_n) \not\in R(A_{\Psi_2})$. By definition of $A_{\Psi_2}$, if $(w_1, \ldots, w_n) \in L(\Psi_2)$, then $(w_1 \# \ldots \# w_n, w_1 \# \ldots \# w_n) \in R(A_{\Psi_2})$. Hence contradiction.

For the other direction of the proof, assume the $n$-PosPTL $S$ is a separator of $L(\Psi_1)$ and $L(\Psi_2)$. Then $S$ can be rewritten as the union of $(L_1 \times L_2 \times \ldots \times L_n)$ where $L_i$s are piece languages. Replace each $n$-piece language $(L_1 \times L_2 \times \ldots \times L_n)$ of $S$ with the $2$-piece language $([L_1 \# L_2 \# \ldots \# L_n] \times ([\Sigma \cup \{\#\})^* \# \ldots \# ([\Sigma \cup \{\#\})^*) \times [\Sigma^* a_1 \Sigma^* \ldots a_n \Sigma^*] \times [\Sigma^* b_1 \Sigma^* \ldots b_m \Sigma^*]$. Denote the union of such languages by $S'$.

It is a $2$-PosPTL over $(\Sigma \cup \{\#\})$. We show that $S'$ is a $2$-PosPTL separator of $R(A_{\Psi_1})$ and $R(A_{\Psi_2})$. Let $(v, u) = (v_1 \# \ldots \# v_n, u_1 \# \ldots \# u_n) \in R(A_{\Psi_1})$, then $(v_1, \ldots, v_n) \in L(\Psi_1)$, and thus $(v_1, \ldots, v_n) \in S$ (since $S$ is a separator). This implies that $(v, u) \in S'$ by its construction. Suppose $(v, u) = (v_1 \# \ldots \# v_n, u_1 \# \ldots \# u_n) \in R(A_{\Psi_2}) \cap S'$, then $(v_1, v_2, \ldots, v_n) \in L(\Psi_2)$. Also $(v_1, v_2, \ldots, v_n) \in S$ by construction of $S'$. This leads to a contradiction that $S$ is separator of $L(\Psi_1)$ and $L(\Psi_2)$. So $S'$ is a $2$-PosPTL separator of $R(A_{\Psi_1})$ and $R(A_{\Psi_2})$.

**Theorem 21.** The $n$-PosPTL separability of right-sided SL formulas is in Expspace and is Pspace-Hard.

**Proof.** Given two right-sided SL formulas $\Psi_1$ and $\Psi_2$, one can construct corresponding two way transducers $A_{\Psi_1}$ and $A_{\Psi_2}$ with polynomial states, as mentioned in Lemma [19].
We conclude with 2 examples. One, we give a string program whose safety checking boils down to checking the separability of two SL string constraints. Second, we illustrate how to compute the PosPTL separator given two SL string constraints, using the 2NFT encoding and Theorem 13.

**Example 22.** As a practical motivation of PosPTL (and PTL), consider the following pseudo-PHP code obtained as a variation of the code at [1]. In this code, a user is prompted to change his password by entering the new password twice.

```pseudo-PHP
define('PASSWORD', 'P. A. Abdulla, M. F. Atig, V. Dave, S. Krishna XX:15

\begin{equation}
\phi\end{equation}

The user inputs the old password \texttt{oldIn} and the new password twice: \texttt{newIn1} and \texttt{newIn2}. These are sanitized and assigned to \texttt{old}, \texttt{new1} and \texttt{new2} respectively. The old sanitized password is compared with the value \texttt{pass} from the database to authenticate the user, and also with the new sanitized password to check that a different password has been chosen, and finally, the sanitized new passwords entered twice are checked to be the same. Sanitization ensures that there are no SQL injections. To ensure the absence of SQL attacks, we require that the query \texttt{query} does not belong to a regular language \texttt{Bad} of bad patterns over some finite alphabet \( \Sigma \) (i.e., the program is safe). This safety condition can be expressed as the unsatisfiability of the following formula \( \varphi \) given by

\begin{equation}
\varphi = T(\texttt{newIn1}) \land \texttt{new2} = T(\texttt{newIn2}) \land \texttt{old} = T(\texttt{oldIn}) \land \texttt{new1} = \texttt{new2} \land \texttt{pass} = \texttt{old} \land \texttt{old} \neq \texttt{new1} \land \texttt{newIn1} = \texttt{newIn2} \land \texttt{query} = u \cdot \texttt{new1} \cdot v \cdot \texttt{userID} \land \texttt{query} \in \texttt{Bad}.
\end{equation}

Note that the check \texttt{new1} = \texttt{new2} has to be done by the server to ensure the sanitized new passwords entered twice are same; however, the check \texttt{newIn1} = \texttt{newIn2} is not redundant, since it can happen that post sanitization, the passwords may agree, but not before. The sanitization on lines 1, 2 and 3 is represented by the transducer \( T \) and \( u, v \) are the constant strings from line 7. It is easy to see that the program given here is safe iff the formula \( \varphi \) is unsatisfiable. Observe that the formula \( \varphi \) is not in the straight line fragment [22] since variable \texttt{new1} has two assignments. Further, it also has a non-benign chain (see below) making it fall out of the fragment of string programs handled in [3]. However the formula \( \varphi \) can be rewritten as a conjunction of the two formula \( \varphi_1 \) and \( \varphi_2 \) in straight-line form where

\begin{align*}
\varphi_1 : &\texttt{new1} = T(\texttt{newIn1}) \land \texttt{old} = T(\texttt{oldIn}) \land \texttt{pass} = \texttt{old} \land \texttt{query} = u \cdot \texttt{new1} \cdot v \cdot \texttt{userID} \land \texttt{query} \in \texttt{Bad} \\
\varphi_2 : &\texttt{new2} = T(\texttt{newIn2}) \land \texttt{new1} = \texttt{new2} \land \texttt{old} \neq \texttt{new1} \land \texttt{newIn1} = \texttt{newIn2}.
\end{align*}
It is easy to see that the program is safe iff $\varphi_1$ and $\varphi_2$ are separable by the PosPTL language that associates to each string variable $\Sigma^*$.

**The string program falls out of the chain-free fragment**

To define the chain-free fragment, [8] introduces the notion of a splitting graph. Assume we are given a string constraint $\Psi = \bigwedge_{i=1}^n \varphi_i$ having $n$ relational constraints, each of the form $\varphi_j = R_j(t_{2j-1}, t_j)$. Let each term $t_i$ be a concatenation of variables $x_{i1} \ldots x_{im}$. Given such a string constraint, its splitting graph contains nodes of the form $\{(i, j) \mid 1 \leq j \leq 2n, 1 \leq i \leq n_j\}$. Each node $(i, j)$ is labeled by the variable $x_{ji}$. The node $(i, 2j - 1)$ (resp. $(i, 2j)$) represents the $i$th term in the left side (respectively, the right side) of constraint $\varphi_j$. There is an edge from node $p$ to node $q$ if there exists a node $p'$ (different from $q$) such that $p$ and $p'$ represent the nodes corresponding to different sides of the same constraint (say $\varphi_i$) and $p'$ and $q$ have the same label. An edge $(p, q)$, with $p = (i, j)$ is labeled by the $j$th constraint $\varphi_j$.

A *chain* in the graph is a sequence of edges of the form $(p_0, p_1)(p_1, p_2) \ldots (p_n, p_0)$. A chain is a *benign* chain if (1) all relational constraints corresponding to the edges are all of the form $R(x, t)$ where $x$ is a single variable $(t$ is a term as usual) and length preserving, and (2) the sequence of positions $p_0, p_1, \ldots, p_n$ all correspond to the left side (or all to the right side).

Recall from $\varphi_1, \varphi_2$ above, that we consider for the string program, the straight-line constraints

\[
\begin{align*}
\phi_1 : \text{new1} &= T(\text{newIn1}), \\
\phi_2 : \text{old} &= T(\text{oldIn}), \\
\phi_3 : \text{pass} &= \text{old}, \\
\phi_4 : \text{new2} &= T(\text{newIn2}), \\
\phi_5 : \text{new1} &= \text{new2}, \\
\phi_6 : \text{old} \neq \text{new1}, \\
\phi_7 : \text{newIn1} &= \text{newIn2}.
\end{align*}
\]

![Figure 1](image.png)

Figure 1 splitting graph for the example

By this definition, the part of the splitting graph which induces a non-benign chain is shown in figure [1]. Observe that there is a chain formed from the nodes of splitting graph which is: $(1, 1) \to (1, 13) \to (1, 8) \to (1, 10) \to (1, 1)$. This chain is not benign since the transducer functions are not length preserving.

**Example 23.** We illustrate how to compute the PosPTL separator given two SL string constraints, using the 2NFT encoding and the result from Theorem [12].

Consider two SL string constraints

\[
\begin{align*}
\Psi_1 : (x, ay) &\in R(T_1) \land y \in (e + f)^*, \\
\Psi_2 : (y, x) &\in R(T_2) \land x \in (e + f)^*.
\end{align*}
\]

The alphabet is $\Sigma = \{a, e, f\}$. The transducer $T_1$ and $T_2$ implement the following functions $\mathcal{T}_1(w) = w \uparrow$ for $w \in a(e + f)^*$, and $\mathcal{T}_2(w') = w' \downarrow$ for $w' \in (e + f)^*$.
Note that $\Psi_1 \land \Psi_2$ is not in $SL$, also it has a non-benign chain as defined in [5]. We show that the languages of formulas $\Psi_1$ and $\Psi_2$ are separable by a $PosPTL$ separator, thereby concluding that $\Psi_1 \land \Psi_2$ is unsatisfiable.

To decide the separability and get this separator, we follow the procedure of Theorem 12.

1. Encode the solutions of formulae to 2NFT.
2. $R(\mathcal{A}_{\Psi_1}) = \{(v#w, u#w) \mid u$ is arbitrary, $w \in (e + f)^*$, and $v \in \{aw\}\}$. 
3. $R(\mathcal{A}_{\Psi_2}) = \{(w#v, w#u) \mid u$ is arbitrary, $w \in (e + f)^*$, and $v \in \{w\}\}$. 
4. Each component in $R(\mathcal{A}_{\Psi_1}), R(\mathcal{A}_{\Psi_2})$ is of the form $\eta(x)\#\eta(y)$. 
5. Decide the emptiness of $R(\mathcal{A}_{\Psi_1}) \uparrow \cap R(\mathcal{A}_{\Psi_2})$. This check reduces to the emptiness of $\min(R(\mathcal{A}_{\Psi_1})) \uparrow \cap R(\mathcal{A}_{\Psi_2})$. The set of minimal words of $R(\mathcal{A}_{\Psi_1}) \uparrow$ is $M = \{(a\# , )\}$ and $M \uparrow \cap R(\mathcal{A}_{\Psi_2})$ is clearly empty. 
6. Hence $M\uparrow = ((\Sigma \cup \{\#\})^*a(\Sigma \cup \{\#\})^* \#(\Sigma \cup \{\#\})^* \times ((\Sigma \cup \{\#\})^* \#(\Sigma \cup \{\#\})^*)$ is a $PosPTL$ separator of $R(\mathcal{A}_{\Psi_1}), R(\mathcal{A}_{\Psi_2})$. This gives the $PosPTL$ separator $\Sigma^*a\Sigma^* \times \Sigma^*$ for $\Psi_1, \Psi_2$.

References

1. PHP Tutorials (2019). URL: https://makephpsites.com/php-tutorials/user-management-tools/changing-passwords.php.
2. Parosh Aziz Abdulla, Mohamed Faouzi Atig, Yu-Fang Chen, Bui Phi Diep, Lukáš Holík, Ahmed Rezine, and Philipp Rümmer. Flatten and conquer: a framework for efficient analysis of string constraints. In PLDI. ACM, 2017.
3. Parosh Aziz Abdulla, Mohamed Faouzi Atig, Yu-Fang Chen, Bui Phi Diep, Lukáš Holík, Ahmed Rezine, and Philipp Rümmer. Trau: SMT solver for string constraints. In FMCAD. IEEE, 2018.
4. Parosh Aziz Abdulla, Mohamed Faouzi Atig, Yu-Fang Chen, Lukáš Holík, Ahmed Rezine, Philipp Rümmer, and Jari Stenman. String constraints for verification. In Computer Aided Verification - 26th International Conference, CAV 2014, Held as Part of the Vienna Summer of Logic, VSL 2014, Vienna, Austria, July 18-22, 2014. Proceedings, volume 8559 of Lecture Notes in Computer Science, pages 150–166. Springer, 2014.
5. Parosh Aziz Abdulla, Mohamed Faouzi Atig, Bui Phi Diep, Lukáš Holík, and Petr Janku. Chain-free string constraints. In Yu-Fang Chen, Chih-Hong Cheng, and Javier Esparza, editors, Automated Technology for Verification and Analysis - 17th International Symposium, ATVA 2019, Taipei, Taiwan, October 28-31, 2019, Proceedings, volume 11781 of Lecture Notes in Computer Science, pages 277–293. Springer, 2019.
6. Parosh Aziz Abdulla and Richard Mayr. Priced timed petri nets. Logical Methods in Computer Science, 9(4), 2013. URL: https://doi.org/10.2168/LMCS-9(4:10)2013 doi:10.2168/LMCS-9(4:10)2013
7. Mohamed Faouzi Atig, Benedikt Bollig, and Peter Habermehl. Emptiness of Ordered Multi-Pushdown Automata is 2ETIME-Complete. Int. J. Found. Comput. Sci., 28(8):945-976, 2017.
8. Félix Baschenis, Olivier Gauvin, Anca Muscholl, and Gabriele Puppis. Minimizing resources of sweeping and streaming string transducers. In 43rd International Colloquium on Automata, Languages, and Programming, ICALP 2016, July 11-15, 2016, Rome, Italy, volume 55 of LIPIcs, pages 114:1-114:14. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2016.
9. Murphy Berzish, Vijay Ganesh, and Yunhui Zheng. Z3str3: A string solver with theory-aware heuristics. In 2017 Formal Methods in Computer Aided Design, FMCAD 2017, Vienna, Austria, October 2-6, 2017, pages 55–59. IEEE, 2017.
10. Murphy Berzish, Yunhui Zheng, and Vijay Ganesh. Z3str3: A string solver with theory-aware branching. CoRR, abs/1704.07935, 2017.
11. Luca Breveglieri, Alessandra Cherubini, Claudio Citrini, and Stefano Crespi-Righizzi. Multi-push-down languages and grammars. Int. J. Found. Comput. Sci., 7(3):253–292, 1996. URL: https://doi.org/10.1142/S0129054196000191 doi:10.1142/S0129054196000191
On the Separability Problem of String Constraints

12 Taolue Chen, Yan Chen, Matthew Hague, Anthony W. Lin, and Zhilin Wu. What is decidable about string constraints with the replaceall function. *PACMPL*, 2(POPL):3:1–3:29, 2018. URL: [https://doi.org/10.1145/3158091](https://doi.org/10.1145/3158091)

13 Taolue Chen, Matthew Hague, Anthony W. Lin, Philipp Rümmer, and Zhilin Wu. Decision procedures for path feasibility of string-manipulating programs with complex operations. *Proc. ACM Program. Lang.*, 3(POPL), January 2019. URL: [http://doi.acm.org/10.1145/3290362](http://doi.acm.org/10.1145/3290362)

14 Taolue Chen, Matthew Hague, Anthony W. Lin, Philipp Rümmer, and Zhilin Wu. Decision procedures for path feasibility of string-manipulating programs with complex operations. *PACMPL*, 3(POPL):49:1–49:30, 2019.

15 Lorenzo Clemente, Wojciech Czerwinski, Slawomir Lasota, and Charles Paperman. Regular separability of parikh automata. In *44th International Colloquium on Automata, Languages, and Programming, ICALP 2017, July 10-14, 2017, Warsaw, Poland*, volume 80 of *LIPIcs*, pages 117:1–117:13. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2017.

16 Bruno Courcelle. On constructing obstruction sets of words. *Bulletin of the EATCS*, 44:178–186, 1991.

17 Wojciech Czerwinski and Slawomir Lasota. Regular separability of one counter automata. *Logical Methods in Computer Science*, 15(2), 2019. URL: [https://lmcs.episciences.org/5563](https://lmcs.episciences.org/5563).

18 Wojciech Czerwinski, Slawomir Lasota, Roland Meyer, Sebastian Muskalla, K. Narayan Kumar, and Prakash Saivasan. Regular separability of well-structured transition systems. In *29th International Conference on Concurrency Theory, CONCUR 2018, September 4-7, 2018, Beijing, China*, volume 118 of *LIPIcs*, pages 35:1–35:18. Schloss Dagstuhl - Leibniz-Zentrum fuer Informatik, 2018.

19 Wojciech Czerwinski, Wim Martens, and Tomás Masopust. Efficient separability of regular languages by subsequences and suffixes. In *Automata, Languages, and Programming - 49th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II*, volume 7966 of *Lecture Notes in Computer Science*, pages 150–161. Springer, 2013.

20 Wojciech Czerwinski, Wim Martens, Lorijn van Rooijen, Marc Zeitoun, and Georg Zetzsche. A characterization for decidable separability by piecewise testable languages. *Discrete Mathematics & Theoretical Computer Science*, 19(4), 2017.

21 Vijay Ganesh and Murphy Berzish. Undecidability of a theory of strings, linear arithmetic over length, and string-number conversion. *CoRR*, abs/1605.09442, 2016. URL: [http://arxiv.org/abs/1605.09442](http://arxiv.org/abs/1605.09442), arXiv:1605.09442.

22 Vijay Ganesh, Mia Minnes, Armando Solar-Lezama, and Martin C. Rinard. Word equations with length constraints: What’s decidable? In *Hardware and Software: Verification and Testing - 8th International Haifa Verification Conference, HVC 2012, Haifa, Israel, November 6-8, 2012, Revised Selected Papers*, volume 7857 of *Lecture Notes in Computer Science*, pages 209–226. Springer, 2012.

23 Vijay Ganesh, Mia Minnes, Armando Solar-Lezama, and Martin C. Rinard. (un)decidability results for word equations with length and regular expression constraints. *CoRR*, abs/1306.6054, 2013. URL: [http://arxiv.org/abs/1306.6054](http://arxiv.org/abs/1306.6054), arXiv:1306.6054.

24 Sheila Ginsburg, Seymour; Greibach. Abstract families of languages. In *Eighth Annual Symposium on Switching and Automata Theory*. IEEE, 1967.

25 G. Higman. Ordering by divisibility in abstract algebras. *Proc. London Math. Soc. (3)*, 2(7), 1952.

26 Lukáš Holík, Petr Janku, Anthony W. Lin, Philipp Rümmer, and Tomáš Vojnar. String constraints with concatenation and transducers solved efficiently. *PACMPL*, 2(POPL):4:1–4:32, 2018. URL: [https://doi.org/10.1145/3158092](https://doi.org/10.1145/3158092), doi:10.1145/3158092.

27 John E. Hopcroft and Jeffrey D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 1979.
Scott Kausler and Elena Sherman. Evaluation of string constraint solvers in the context of symbolic execution. In *ASE ’14*. ACM, 2014.

Adam Kiezun, Vijay Ganesh, Philip J. Guo, Pieter Hooimeijer, and Michael D. Ernst. HAMPI: a solver for string constraints. In *Proceedings of the Eighteenth International Symposium on Software Testing and Analysis, ISSTA 2009*, Chicago, IL, USA, July 19-23, 2009, pages 105–116. ACM, 2009.

Tianyi Liang, Andrew Reynolds, Cesare Tinelli, Clark Barrett, and Morgan Deters. A DPLL(T) theory solver for a theory of strings and regular expressions. In *CAV’14*, volume 8559 of LNCS. Springer, 2014.

Tianyi Liang, Andrew Reynolds, Nestan Tsiskaridze, Cesare Tinelli, Clark W. Barrett, and Morgan Deters. An efficient SMT solver for string constraints. *Formal Methods in System Design*, 48(3):206–234, 2016. URL: https://doi.org/10.1007/s10703-016-0247-6.

Anthony Widjaja Lin and Pablo Barceló. String solving with word equations and transducers: towards a logic for analysing mutation XSS. In *Proceedings of the 43rd Annual ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages, POPL 2016*, St. Petersburg, FL, USA, January 20 - 22, 2016, pages 123–136. ACM, 2016.

Kenneth L. McMillan. Interpolation and sat-based model checking. In *Computer Aided Verification, 15th International Conference, CAV 2003*, Boulder, CO, USA, July 8-12, 2003, *Proceedings*, volume 2725 of Lecture Notes in Computer Science, pages 1–13. Springer, 2003.

Kenneth L. McMillan. An interpolating theorem prover. In *Tools and Algorithms for the Construction and Analysis of Systems, 10th International Conference, TACAS 2004*, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2004, Barcelona, Spain, March 29 - April 2, 2004, *Proceedings*, volume 2988 of Lecture Notes in Computer Science, pages 16–30. Springer, 2004.

Kenneth L. McMillan. Lazy abstraction with interpolants. In *Computer Aided Verification, 18th International Conference, CAV 2006*, Seattle, WA, USA, August 17-20, 2006, *Proceedings*, volume 4144 of Lecture Notes in Computer Science, pages 123–136. Springer, 2006.

Christophe Morvan. On rational graphs. In *Foundations of Software Science and Computation Structures*, Berlin, Heidelberg, 2000. Springer Berlin Heidelberg.

Thomas Place, Lorijn van Rooijen, and Marc Zeitoun. Separating regular languages by piecewise testable and unambiguous languages. In *Mathematical Foundations of Computer Science 2013 - 38th International Symposium, MFCS 2013*, Klosterneuburg, Austria, August 26-30, 2013. *Proceedings*, volume 8087 of Lecture Notes in Computer Science, pages 729–740. Springer, 2013.

 Wojciech Plandowski. An efficient algorithm for solving word equations. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC ’06, pages 467–476, New York, NY, USA, 2006. ACM. URL: http://doi.acm.org/10.1145/1132516.1132584.

Andrew Reynolds, Maverick Woo, Clark W. Barrett, David Brumley, Tianyi Liang, and Cesare Tinelli. Scaling up DPLL(T) string solvers using context-dependent simplification. In *Computer Aided Verification - 29th International Conference, CAV 2017*, Heidelberg, Germany, July 24-28, 2017, *Proceedings, Part II*, volume 10427 of Lecture Notes in Computer Science, pages 453–474. Springer, 2017.

Arto Salomaa. Formal languages. In *Formal Languages*. Academic Press, 1987.

Prateek Saxena, Devdatta Akhawe, Steve Hanna, Feng Mao, Stephen McCamant, and Dawn Song. A symbolic execution framework for javascript. In *31st IEEE Symposium on Security and Privacy, SP’10*, 16-19 May 2010, Berkeley/Oakland, California, USA, pages 513–528. IEEE Computer Society, 2010.

Prateek Saxena, Steve Hanna, Pongsin Poosankam, and Dawn Song. FLAX: systematic discovery of client-side validation vulnerabilities in rich web applications. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2010*, San Diego, California, USA, 28th February - 3rd March 2010. The Internet Society, 2010.
On the Separability Problem of String Constraints

43 Imre Simon. Piecewise testable events. In Automata Theory and Formal Languages, 2nd GI Conference, Kaiserslautern, May 20-23, 1975, volume 33 of Lecture Notes in Computer Science, pages 214–222. Springer, 1975.

44 Minh-Thai Trinh, Duc-Hiep Chu, and Joxan Jaffar. S3: A symbolic string solver for vulnerability detection in web applications. In CCS’14. ACM, 2014.

45 Minh-Thai Trinh, Duc-Hiep Chu, and Joxan Jaffar. Progressive reasoning over recursively-defined strings. In CAV’16, volume 9779 of LNCS. Springer, 2016.

46 Moshe Y. Vardi. A note on the reduction of two-way automata to one-way automata. Inf. Process. Lett., 30(5):261–264, 1989. URL: https://doi.org/10.1016/0020-0190(89)90205-6, doi:10.1016/0020-0190(89)90205-6.

47 Hung-En Wang, Tzung-Lin Tsai, Chun-Han Lin, Fang Yu, and Jie-Hong R. Jiang. String analysis via automata manipulation with logic circuit representation. In Computer Aided Verification - 28th International Conference, CAV 2016, Toronto, ON, Canada, July 17-23, 2016, Proceedings, Part I, volume 9779 of Lecture Notes in Computer Science, pages 241–260. Springer, 2016.

48 Fang Yu, Muath Alkhalaf, and Tevfik Bultan. Stranger: An automata-based string analysis tool for PHP. In Tools and Algorithms for the Construction and Analysis of Systems, 16th International Conference, TACAS 2010, Held as Part of the Joint European Conferences on Theory and Practice of Software, ETAPS 2010, Paphos, Cyprus, March 20-28, 2010. Proceedings, volume 6015 of Lecture Notes in Computer Science, pages 154–157. Springer, 2010.

49 Yunhui Zheng, Xiangyu Zhang, and Vijay Ganesh. Z3-str: A Z3-based string solver for web application analysis. In ESEC/FSE’13. ACM, 2013.
Appendix

A Proof from Section 2 (Proof of lemma 1)

Given a string constraint $\Psi$ over variables $x_1, \ldots, x_m$, one can construct two SL string constraints $\Psi_1$ and $\Psi_2$ s.t. $\Psi$ and $\Psi_1 \land \Psi_2$ are equisatisfiable.

By definition $\Psi$ is a conjunction of atomic regular membership and relational string constraints i.e. $\Psi = \bigwedge_{i=1}^{n} \phi_i \land \bigwedge_{i=1}^{m} \varphi_i$, where $\phi_i$s are membership constraints and $\varphi_i$s are relational constraints. Since regular languages are closed under intersection, one can combine the constraints of the form $x_i \in L_1$ and $x_i \in L_2$ as $x_i \in L$ where $L = L_1 \cap L_2$. This way we have at most one membership constraint for each variable. We can add the conjunction of membership constraints to either $\Psi_1$ or $\Psi_2$ or both.

Next, we show that by introducing new variables corresponding to each relational constraint $\varphi$, we can partition $\Psi$ into two SLs. For each relational constraint $\varphi: (x_i, t_i) \in R(T)$, we add a new variable $u_\varphi$ and add a formula $u_\varphi = x_i$ (or $(u_\varphi, t_i) \in Id$, where $Id$ is the identity) to $\Psi_1$ and $(u_\varphi, t_i) \in R(T)$ to $\Psi_2$. After iterating this procedure for all relational constraints $\varphi$, we claim that we have SL formulas in $\Psi_1$ as well as $\Psi_2$.

Observe that in $\Psi_i$ for $i = 1, 2$, variables $x_1, \ldots, x_m$ appear only as part of (i) membership constraint or (ii) input to relational constraint. These variables are independent. Newly introduced variables $u_{\varphi_j}$ appear as only output of exactly one relational constraint in $\Psi_1$ and $\Psi_2$. Hence the variables can be ordered as $u_{\varphi_j}$ having lower precedence than $x_i$s, and both formulas $\Psi_1$ and $\Psi_2$ preserve the SL syntax. Clearly $\Psi$ and $\Psi_1 \land \Psi_2$ are equisatisfiable: If we have a solution for $x_i$s in $\Psi$, then the same assignment works in $\Psi_1 \land \Psi_2$, in addition $u_{\varphi_j}$s are assigned the values based on values of $x_i$s. On other side, given a solution for $\Psi_1 \land \Psi_2$, the same assignment of variables $x_i$s works for $\Psi$.

B Proof from Section 3 (Proof of theorem 2)

In the following, we show that the REG separability problem is undecidable even for the subclass SL of string constraints.

Proof. The proof is done by reduction from the halting problem of Turing machines. Consider a deterministic Turing machine $M$ whose set of states is $Q$ and tape alphabet is $\Sigma$, with $a, \#, \not\in \Sigma$. Encode a configuration $w$ of $M$ as a word in $\Sigma^* \times (Q \times \Sigma^*) \times \Sigma^*$ and use $w \rightarrow_M w'$ to denote that $M$ can reach the configuration $w'$ from the configuration $w$ in one step. Let $w_1$ denote the initial configuration of $M$. It can be easily seen that one can design a transducer that accepts the language $\{(w, w') \mid w \rightarrow_M w'\}$. This construction can be extended to design two transducers $\mathcal{T}_1$ and $\mathcal{T}_2$ accepting respectively the following two languages:

$L_1 = \{(w_1 \# \# w_2 \# \cdots \# w_{2k-1} \# a^j \# w_2 \# w_4 \# \cdots \# w_{2k} \# a^j \# \cdots \# w'_j \# \mid k \geq 1, w_{2i-1} \rightarrow_M w_{2i} \text{ for } i \in [1, k], \text{ and } w'_1, \ldots, w'_j \in \Sigma^* \times (Q \times \Sigma) \times \Sigma^* \text{ are arbitrary}\}$

$L_2 = \{(w_1 \# \# w_2 \# \cdots \# w_{2k-1} \# a^j \# w_2 \# w_4 \# \cdots \# w_{2k} \# a^j \# \cdots \# w'_j \# \mid k \geq 1, w_{2i-2} = w_{2i-1} \text{ for } i \in [2, k], \text{ and } w'_1, \ldots, w'_j \in \Sigma^* \times (Q \times \Sigma) \times \Sigma^* \text{ are arbitrary}\}$

Consider the SL formulae $\Psi_1 = (x, y \# y) \in \mathcal{R}(\mathcal{T}_1)$ and $\Psi_2 = (x, y \# y) \in \mathcal{R}(\mathcal{T}_2)$. Then, it is easy to see that

$L(\Psi_1) = \{(w_1 \# \# w_2 \# \cdots \# w_{2k-1} \# a^k \# w_2 \# w_4 \# \cdots \# w_{2k} \# \mid k \geq 1, w_2, w_3, \ldots, w_{2k} \in \Sigma^* \times (Q \times \Sigma) \times \Sigma^*, \text{ and } w_{2i-1} \rightarrow_M w_{2i} \text{ for } i \in [1, k]\}$
On the Separability Problem of String Constraints

We construct the which checks if the number of configurations in \((\Psi, 2k)\times (\Omega \times \Sigma^*) \times \Sigma^*, \omega_{2k-2} = \omega_{2k-1} \text{ for } i \in [2, k]\)

We prove that \(\Psi_1\) and \(\Psi_2\) are REG separable if and only if \(M\) halts.

**Case 1.** Suppose \(M\) does not halt. Assume that there exists a REG separator \(\bigcup_{i=1}^n L_i \times L'_i\) separating languages of \(\Psi_1\) and \(\Psi_2\). Let \(A_i, B_i\) respectively be DFA s.t. \(L_i = L(A_i), L'_i = L(B_i)\). Let \(n-1\) be the total number of states in \(A_1, \ldots, A_t\). Consider the 2-tuple

\[((v_1, v_2) = (w_1 \# w_2 \# \ldots \# w_i \# \$ a^i \# w_{i+1} \ldots \# w_n \# \$ w_{i+2} \ldots \# w_{n+1} \ldots \# w_{n+2} \ldots \# w_{n+k} \ldots \# w_{n+k})\]

where \(u_i \rightarrow_M u_{i+1}\) for all \(i \in [2, n]!\) and \(w_1 \rightarrow_M w_2\). Then \((v_1, a^i, v_2) \in L(\Psi_1)\) and \((v_1, a^2, a^3, v_2) \in L(\Psi_2)\). Suppose that \((v_1, a^i, v_2) \in L(A_i)\times L(B_i)\) and that \(\delta_{A_i}(q_0, v_1) = p_i\) and \(\delta_{A_i}(p_i, a^i) = r_i\) (where \(\delta_{A_i}\) denotes the transition function of \(A_i\) and \(q_0\) is the initial state of \(A_i\)). Then by the pigeon hole principle on the number of states, we must encounter a loop and hence repeat at least one state while reading \(a^i\) from \(p_i\) in \(A_i\). Let \(s_i\) be the period (length of the loop) \(1 \leq s_i \leq n\) such that \(\delta_{A_i}(p_i, a^{n+j} s_i) = r_i\) for any \(j \in \mathbb{N}\). In particular, if we choose \(j = n! \div s_i\), we get \(\delta_{A_i}(p_i, a^{n} s_i) = r_i\). Thus, \((v_1, a^{n} s_i, v_2)\) is also in \(\bigcup_{i=1}^n L_i \times L'_i\) and which is a contradiction.

**Case 2.** Suppose \(M\) halts in \(n > 1\) steps. Then \((w_1 = u_1) \rightarrow_M u_2 \rightarrow_M \ldots \rightarrow_M u_n \rightarrow_M u_{n+1}\) and constructs the REG separator for \(L(\Psi_1)\) and \(L(\Psi_2)\) as follows. Let \(\text{ config}\) represent any configuration of \(M\), and let \(R_1 = (\text{ config } \#)^* \text{ config } \# a^*\), \(R_2 = (\text{ config } \#)^* \text{ config}\). Then \(R_1, R_2\) are regular languages.

We first consider the case when we have \((u, v) \in R_1 \times R_2\) s.t. number of configurations in both \(u, v\) is less than \(n\). We construct \(L_i\) which checks whether the number of configurations in \(u\) is \(i\), and the number of \(a\)'s in \(u\) is also equal to \(i\). Likewise, we construct language \(L_i\) which checks if the number of configurations in \(v\) is \(i\). Then \(\bigcup_{i=1}^n L_i \times L'_i\) does not intersect with \(L(\Psi_2)\) and contains \(L(\Psi_1)\). Thus for each \(i \in [1, n]\),

\[L_i = \{w_1 \# w_2 \# \ldots \# w_{2i} \# a^i \mid w_1, w_3, \ldots, w_{2i-1} \in \Sigma^* \times (Q \times \Sigma) \times \Sigma^*\}\]

\[L'_i = \{w_2 \# w_4 \# \ldots \# w_{2i} \mid w_2, w_4, \ldots, w_{2i} \in \Sigma^* \times (Q \times \Sigma) \times \Sigma^*\}\]

Now consider the case when the number of configurations is more than \(n\) in \(u\) or \(v\), for \((u, v) \in R_1 \times R_2\). It suffices to choose non-deterministically an index \(j \leq n\) and check if \(w_{2j}\) and \(w_{2j+1}\) are not equal and all the configurations before \(w_{2j+1}\) and \(w_{2j}\) follow the run of \(M\). It is sufficient to check up to \(n\) configurations since the run in \(M\) has only \(n\) configurations in the sequence. For \(1 \leq j \leq n\), consider the subset \(R_j\) of \(R_1 \times R_2\) defined as \(R_j = \{u \# u_{j} \# w_{j-1} \# w_{j+1} \# \# \mid u \in L(\Psi_1)\\times (Q \times \Sigma) \times \Sigma^*\} \times \{u \# u_{j} \# w_{j+1} \# \# \mid u \in L(\Psi_1)\times (Q \times \Sigma) \times \Sigma^*\}\) such that \(u_j \neq w_{2j+1}\). \(R_j\) is regular since the length of configurations \(u_j\) and \(w_{2j+1}\) are bounded. \(\bigcup_{j=1}^n R_j \cap L(\Psi_2) = \emptyset\).

Now consider a pair \((u, v) \in R_1 \times R_2\), where there is no violation as described above in the equality between \(u_j\) and \(w_{2j+1}\) until \(j = n\). Then we cannot continue finding a successor for \(u_n\) as a result of which, such pairs \((u, v) \notin L(\Psi_1)\), and hence cannot be considered in separator.

This way we remove all words in \(L(\Psi_2)\) and accept all the words in \(L(\Psi_1)\). Hence \((\bigcup_{i=1}^n L_i \times L'_i) \cup (\bigcup_{j=1}^n R_j)\) is a REG separator.
C Proofs from Section 4

C.1 Proof of Lemma 3

We show that the class of ordered multi pushdown languages (OMPL) form a full trio.

Proof. Let $L$ be an OMPL over alphabet $\Sigma$ and let $A$ be the OMPA accepting $L$. We now show the effective closure wrt the three properties.

• For $B \subseteq \Sigma$, construct an OMPA from $A$ by replacing transitions on input symbols $a \in \Sigma \setminus B$ with $\epsilon$. The resulting OMPA accepts the $B$-projection of $L$. For $B \not\subseteq \Sigma$, the OMPA accepting the emptyset suffices.

• For the $B$-upward closure of $L$, add extra transitions to $A$ as loops on each state on any input symbol from $B$, with no push/pop operations. The resultant OMPA will accept the $B$-upward closure of $L$.

• The third property is the closure wrt intersection of regular languages. It is well-known that OMPL are effectively closed wrt intersection with regular languages. This shows that OMPLs form a full trio.

C.2 Proof of Lemma 5

In the following, we present the proof to show that each OMPL $L$ has a semilinear Parikh image.

Proof. To see this, first note that using the result of [7], one can construct a $D_n$-grammar equivalent to the OMPL. (We refer the reader to [7] for the definition of $D_n$-grammar.) Next, using the result from [11], we know that for each $D_n$-grammar $G$ there is an underlying context-free grammar $G'$ s.t. for each $w \in L(G)$, there is a $w' \in L(G')$ (and conversely) such that $w$ and $w'$ have the same Parikh image. Thus, the language generated by the $D_n$-grammar $G$ and the language generated by the context-free grammar $G'$ have the same Parikh image. Since the Parikh image of $L(G')$ is semilinear [40] we obtain that the language of each $D_n$-grammar is semilinear, giving us the result (since $D_n$-grammars are equivalent to the OMPL).

C.3 Details for Lemma 8

Proof. For intuition of this proof, readers can refer the proof provided in Lemma 8. Formally, set of states of the OMPA $A_i$ is states of $T_i$ in addition with the states $\{q_i^{\text{init}}, q_i^1, q_i^2, q_i^3, q_i^{\text{final}}\}$ to distinguish phase 1, 2, and 4. Initial state is defined as $q_i^{\text{init}}$.

Phase 1: Being in state $q_i^{\text{init}}$, content of first $(n - i)$ stacks are moved to next $(n - i)$ stacks i.e. content of stack $j$, is moved to stack $n - i + j$ for all $1 \leq j \leq n - i$. To achieve this, for all $j$, $1 \leq j \leq n - i$, for all $a \in \Sigma$, we have transition $(q_i^{\text{init}}, \bot ((j - 1) \text{ times}), a, \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (q_i^{\text{init}}, \epsilon((n - i + j - 1) \text{ times}), a, \epsilon, \ldots, \epsilon)$. 

![Diagram](image-url)
Phase 2: After phase 1, stacks indexed 1 to \( n - i \) are empty and next \( n - i \) stacks contain the valuation of variables \( x_{i+1} \) to \( x_n \) in reverse. To distinguish this phase, we add a no operation transition from \( q_i^{init} \) to \( q_i^2 \). For this, we have a transition

\[
(q_i^{init}, \perp ((n - i) \text{ times}), \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (q_i^2, \epsilon, \ldots, \epsilon)
\]

In this phase, content of non-empty stacks \( (n - i + 1), \ldots, (2n - 2i) \) is moved to last \( (n - i) \) stacks indexed \( (2n - 2i + \vert t_i \vert + 2) + 1, \ldots, (2n - 2i + \vert t_i \vert + 2) + (n - i) \). Also, if any of the variable from \( x_{i+1}, \ldots, x_n \) is present in \( t_i \) at position \( j \) then the content of the corresponding stack is pushed to the stack \( (2n - 2i) + j \). It is done with the following transitions. For all \( k, 1 \leq k \leq (n - i) \), for all \( a \in \Sigma \), we have

\[
(q_i^2, \perp ((n - i + k - 1) \text{ times}), a, \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (q_i^2, a((2n - 2i) \text{ times}), \epsilon, \ldots, \epsilon),
\]

\[
(\text{a if } t_i[1] = x_{i+k}; \text{ else } \epsilon, \epsilon, \ldots, \epsilon),
\]

\[
(\text{a if } t_i[2] = x_{i+k}; \text{ else } \epsilon, \epsilon, \ldots, \epsilon),
\]

\[
(\ldots, \ldots, \ldots, \epsilon, \epsilon).
\]

Phase 3: After phase 2, we have valuation of \( t_i[j] \) present in stack \( (2n - 2i + j) \). Also, last \( (n - i) \) stacks indexed \( (2n - 2i + \vert t_i \vert + 3) \) to \( (3n - 3i + \vert t_i \vert + 2) \) contains the valuation of variables \( x_{i+1}, \ldots, x_n \) respectively. In this phase, we evaluate the valuation of \( x_i \) based on valuation of \( t_i, \eta(t_i) \). Observe that \( \eta(t_i) \) is stored in the sequence in stacks \( (2n - 2i + 1), \ldots, (2n - 2i) + \vert t_i \vert \) and all the stacks before these, are empty. Thanks to OMPA, one can pop the content of these stacks one after another, and find the output of these words produced by \( \mathcal{T} \). Output produced is pushed into the stack indexed \( (2n - 2i + \vert t_i \vert + 1) \). To begin this process, we use a no operation transition from \( q_i^2 \) to \( p_0 \) where \( p_0 \) belongs to initial states of \( \mathcal{T}_i \):

\[
(q_i^2, \perp ((2n - 2i) \text{ times}), \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (p_0, \epsilon, \ldots, \epsilon)
\]

where \( p_0 \) is in set of initial state of \( \mathcal{T}_i \). To simulate \( \mathcal{T}_i \) on content of these \( \vert t_i \vert \) stacks, for every transition \((p, (b, a), p') \) of \( \mathcal{T}_i \), and for all \( j \) s.t. \( 1 \leq j \leq \vert t_i \vert \) we have following corresponding transitions in \( \mathcal{A}_i \):

\[
(p, \perp (2n - 2i + (j - 1) \text{ times }), a, \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (p', \epsilon((2n - 2i + \vert t_i \vert) \text{ times }), b, \epsilon, \ldots, \epsilon)
\]

Phase 4: After finishing computing value of \( x_i \), if the state of \( \mathcal{T}_i \) reached is a final state, \( \mathcal{A}_i \) moves to state \( q_i^3 \) to start the process of this phase.

\[
(p_f, \perp (2n - 2i + \vert t_i \vert \text{ times }), \epsilon, \ldots, \epsilon) \rightarrow^\epsilon (q_i^3, \epsilon, \ldots, \epsilon)
\]
We have already computed the valuation of variable $x_i$ in previous phase. But notice that it is stored in stack $(2n - 2i + |t_i| + 1)$ in reverse order i.e. the top of the stack is the last letter of value of $x_i$, while the bottom represents the first letter of value of $x_i$. So there is a next stack indexed $(2n - 2i + |t_i| + 2)$ reserved to store its reverse. The operation is carried out using following transition, for all $a \in \Sigma$:

$$(q_i^4, \bot(2n - 2i + |t_i|) \text{ times }, a, \epsilon, \ldots, \epsilon) \rightarrow^* (q_i^4, \epsilon(2n - 2i + |t_i| + 1 \text{ times }), a, \epsilon, \ldots, \epsilon)$$

This operation is applied until the stack $(2n - 2i + |t_i| + 2)$ becomes empty. Thus at the end of this phase, we have valuation of $x_i, \ldots, x_n$ in last $(n - i + 1)$ stacks i.e. indexed $(2n - 2k + |t_i| + 2), (3n - 3k + |t_i| + 2)$, and $A_i$ moves to accepting state $q_i^{final}$.

$$(q_i^4, \bot(2n - 2i + |t_i| + 1 \text{ times }), \epsilon, \ldots, \epsilon) \rightarrow^* (q_i^{final}, \epsilon, \ldots, \epsilon)$$

End of phase 4

C.4 Proof of Lemma 10

Let $\Psi_1$ and $\Psi_2$ be two SL formulae with $x_1, \ldots, x_n$ as their set of variables. Let $A_1$ and $A_2$ be two OMPAs such that $L(A_1) = \text{Encode}(L(\Psi_1))$ and $L(A_2) = \text{Encode}(L(\Psi_2))$. $\Psi_1, \Psi_2$ are $n$-PTL separable iff $A_1, A_2$ are PTL-separable.

Proof. Lemma 9 gives OMPAs $A_{\Psi_1}$ and $A_{\Psi_2}$ s.t. $L(A_{\Psi_1}) = \text{Encode}(L(\Psi_1))$. Let $L$ be a PTL s.t. $L(A_{\Psi_1}) \subseteq L$ and $L(A_{\Psi_2}) \cap L = \emptyset$. Both $L(A_{\Psi_1})$ and $L(A_{\Psi_2})$ are sets of words of the form $\{w_1w_2\ldots w_n | w_i \in \Sigma^*\}$, and hence the PTL $L$ is a finite Boolean combination of languages from $(\Sigma^* \# ^\ast)$. Let $S \subseteq (\Sigma^* \# ^\ast)$ be a PTL consisting of all words having exactly $n - 1$ #’s. $S$ is a finite Boolean combination of piece languages $L_i$ (assume $1 \leq i \leq p$ for some $p \in \mathbb{N}$) where each $L_i$ has the form $L_{i1} \# L_{i2} \# \ldots \# L_{in}$. That is, $L_{i1}, \ldots, L_{in}$ are all piece languages over $\Sigma$. It suffices to consider $L = S$.

Given the PTL $S$ over $\Sigma \cup \{\#\}$, define the language $S'$ over $\Sigma^n$, as a finite Boolean combination of languages $L_{i1} \times \ldots \times L_{in}$ for $1 \leq i \leq p$. Since $L_{i1}, \ldots, L_{in}$ are piece languages, $S'$ is a PTL over $\Sigma^n$. We show that $S'$ separates $L(\Psi_1)$ and $L(\Psi_2)$. Consider a word $(w_1, \ldots, w_n) \in S'$. Then indeed, $w_1 \# w_2 \# \ldots \# w_n \in S$ by definition. Since $S$ separates $L(A_{\Psi_1})$ and $L(A_{\Psi_2})$, we know that $w_1 \# w_2 \# \ldots \# w_n \in L(A_{\Psi_1})$, but not in $L(A_{\Psi_2})$. By definition of $\text{Encode}$, this gives $(w_1, \ldots, w_n) \in L(\Psi_1)$, but not in $L(\Psi_2)$. Since $S'$ is a PTL over the $n$-tuple alphabet, it is an $n$-PTL separating $L(\Psi_1)$ and $L(\Psi_2)$.

The converse follows on similar lines.

D.1 Proof of Theorem 12

Two languages $I$ and $E$ from a class $C$ are $n$-PosPTL separable iff $I |\cap E = \emptyset$ iff $I \cap E_\downarrow = \emptyset$. 

D.2 Proofs from Section 5
Proof. The proof is done for the class of PosPTL however it can easily be extended to languages over n-tuples words.

Let $I, E$ be languages over a finite alphabet $\Sigma$. Assume $I$ and $E$ are separable by the PosPTL $S$. Then $I \subseteq S$ and $S \cap E = \emptyset$. Since $S$ is upward closed, $I^\uparrow \subseteq S$ and hence $I^\uparrow \cap E = \emptyset$.

Assume $I^\uparrow \cap E = \emptyset$. We claim that $I^\uparrow$ is a PosPTL-separator. We have already that $I \subseteq I^\uparrow$ and $I^\uparrow \cap E = \emptyset$. Now we will show that $I^\uparrow$ is PosPTL: $\text{min}(I^\uparrow)$ is a finite set since the subword ordering $\preceq$ is a wqo on $\Sigma^*$. Let $\text{min}(I^\uparrow) = \{w_1, \ldots, w_k\}$. Then $I^\uparrow$ can be defined as a finite union of $\{w_i\}^\uparrow$ for $i \in \{1, 2, \ldots, k\}$. For $w_i = a_{i_1} \ldots a_{i_m}$, with $a_{i_j} \in \Sigma$, $\{w_i\}^\uparrow$ is the piece language $\Sigma^*a_{i_1}\Sigma^* \ldots \Sigma^*a_{i_m}\Sigma^*$. Thus, $I^\uparrow$ is indeed in PosPTL, being the finite union of piece languages.

To see $I^\uparrow \cap E = \emptyset$ iff $I \cap E \downarrow = \emptyset$. Let $w \in I^\uparrow \cap E$. Let $w' \preceq w$ be s.t. $w' \in I$. Then $w' \in I \cap E \downarrow$. The converse works similarly. □

D.2 PosPTL separability of PDA (Proof of Theorem 14)

We show that PosPTL separability of context free languages is PSPACE-COMPLETE. Context free languages are equivalently represented by pushdown automata(PDA) or context free grammars(CFG) and the conversion from CFG to PDA and PDA to CFG is in polynomial time. We recall the definition of context free grammar briefly here.

A Context Free Grammar is defined by a tuple $G = (N, T, R, S)$ where $N$ and $T$ are finite set of non-terminals and terminals respectively, $S$ is an initial non-terminal. $R$ is a finite set of production rules of the form (i) $A \rightarrow BC$, (ii) $A \rightarrow a$, or (iii) $S \rightarrow \epsilon$ where $A \in N$, $B, C \in N \setminus \{S\}$ and $a \in T$. We define derivation relation $\Rightarrow$ as following: Given $u_1, u_2 \in (N \cup T)^*$, $u_1 \Rightarrow u_2$ iff there exists $A \rightarrow w$ in the production rules set $R$ such that $u_1 = vAu'$ and $u_2 = uvwv'$ for some $v, v' \in (N \cup T)^*$. Let $\Rightarrow^*$ denote one or more application of derivation relation $\Rightarrow$. A string $w$ belongs to the language generated by grammar $G$ iff it can be generated starting from $S$, applying derivation rules one or more times. Hence, the language generated by the grammar $G$ is $L(G) = \{w \in T^* | S \Rightarrow^* w\}$.

Proof of Theorem 14. We show that PosPTL separability problem of PDA is PSPACE-COMPLETE. Given PDA $A_1$ and $A_2$, consider their corresponding CFGs $G_1, G_2$. $L(G_1)^\uparrow \cap L(G_2)^\downarrow \neq \emptyset$ iff $L(G_1)^\uparrow \cap L(G_2)^\downarrow \neq \emptyset$. If there is a witness of $L(G_1)^\uparrow \cap L(G_2)^\downarrow$, the same witness trivially works for $L(G_1)^\uparrow \cap L(G_2)^\downarrow$. Now suppose $w \in L(G_1)^\uparrow \cap L(G_2)^\downarrow$, then there exists $w' \in L(G_2)$ s.t. $w \preceq w'$. Also, $w' \in L(G_1)^\uparrow$. Hence, we have a witness of the non-emptyness of $L(G_1)^\uparrow \cap L(G_2)^\downarrow$. We prove that the non-emptyness of $L(G_1)^\uparrow \cap L(G_2)^\downarrow$ is PSPACE-complete.

PSPACE-membership. The non-emptyness check of $L(G_1)^\uparrow \cap L(G_2)^\downarrow$ can be seen to be in PSPACE as follows. We construct PDAs $A_{G_1}, A_{G_2}$ respectively for $L(G_1)^\uparrow$ and $L(G_2)^\downarrow$. This construction takes polynomial time; moreover, $A_{G_1}, A_{G_2}$ use only a bounded stack which is polynomial in the size of the CFG ($G_1$ or $G_2$). This fact is rather easy to see for $A_{G_1}$ (Lemma 21), but is quite involved for $A_{G_2}$ (Lemma 25). Assuming the construction of $A_{G_1}, A_{G_2}$, we give an NPSACE algorithm as follows. Guess a word $w$, one symbol at a time, and run $A_{G_1}, A_{G_2}$ in parallel. Since both the PDAs require only a polynomially bounded stack size, we need polynomial space to store the information pertaining to the states, stacks and the current input symbol. If $A_{G_1}, A_{G_2}$ accept $w$, we are done. The PSPACE-membership follows from Savitch’s theorem.

PSPACE-hardness. We reduce the halting problem of a Linear Bounded Turing Machine (LBTM) to our problem. Given a LBTM $M$ and a word $w$, we construct PDAs $P_1$ and $P_2$.
such that \( \mathcal{L}(P_1)^\uparrow \cap \mathcal{L}(P_2)^\downarrow \neq \emptyset \) iff \( M \) accepts \( w \). Our construction is in Appendix D.3.

**Bounded Stack Size of \( A_{G\uparrow} \), \( A_{G\downarrow} \)**

We first consider the case of \( A_{G\uparrow} \). The proof follows by examining the height of derivation trees of words in \( \min(\mathcal{L}(G)) \), and showing that they have a bounded height. This bound gives the height of the stack size in \( A_{G\uparrow} \). The proof of Lemma 24 is in Section D.6.

\[ \blacktriangleright \textbf{Lemma 24.} \quad \text{Given a CFG } G = (N, T, R, S), \text{ the PDA } A_{G\uparrow} \text{ can be constructed in polynomial time, using a stack whose height is polynomial in the size of } G. \quad \mathcal{L}(A_{G\uparrow}) = \mathcal{L}(G)^\uparrow. \]

\[ \blacktriangleright \textbf{Lemma 25.} \quad \text{Given CFG } G = (N, T, R, S), \text{ one can construct PDA } A_{G\downarrow} \text{ in polynomial time which uses polynomially bounded stack for computation such that } \mathcal{L}(A_{G\downarrow}) = \mathcal{L}(G)^\downarrow. \]

**Proof.** First, we compute a CFG \( G' \) using \( G \), and then construct \( A_{G\downarrow} \) using \( G, G' \). The construction of \( G' \) is described below. We recall the approach described in [16] to compute downward closure, where a regular expression is computed for \( \mathcal{L}(G)^\downarrow \).

For every language \( L \), let \( \alpha(L) \) be the set of terminal symbols occurring in \( L \) (hence \( \alpha(L) = \emptyset \) iff \( L \subseteq \{ \} \)). For \( L', L'' \subseteq T^* \) we have: (i) \( \alpha(L \cup L') = \alpha(L) \cup \alpha(L') \), (ii) \( (L \cup L'') \downarrow = L \downarrow \cup L'' \downarrow \), and (iii) \( LL' \downarrow = L \downarrow L' \downarrow \). For \( m \in (N \cup T)^* \), let \( L(G, m) \) denote the language generated by \( G \), starting from the word \( m \). For every \( A, B \in N \), define:

1. \( B <_1 A \) iff \( A \Rightarrow_G^* m B m' \) for some \( m, m' \in (N \cup T)^* \).
2. \( B <_2 A \) iff \( A \Rightarrow_G^* m B m' B m'' \) for some \( m, m', m'' \in (N \cup T)^* \).
3. \( B =_1 A \) iff \( B <_1 A <_1 B \).

A proof of the first implication in Claim 26 is in Appendix D.3, the second implication has a similar proof.

\[ \blacktriangleright \textbf{Claim 26.} \quad A <_2 A \Rightarrow L(G, A)^\downarrow = (\alpha(L(G, A)))^\downarrow, \quad A =_1 B \Rightarrow L(G, A)^\downarrow = L(G, B)^\downarrow. \]

**Computing \( L(G, A)^\downarrow \).** Now, we explain how \( L(G, A)^\downarrow \) can be computed for any given \( A \in N \). If \( A <_2 A \), then claim 26 yields the answer. Otherwise, we compute \( L(G, A)^\downarrow \) in terms of \( L(G, B)^\downarrow \) for \( B <_1 A \) and \( B \neq A \), assuming \( L(G, B)^\downarrow \) is given by previously computed rules for \( L(G, B) \). Let \( p : A \rightarrow m \) be a production rule. We define the words \( R_0(p), R_1(p) \) and \( R_2(p) \) as follows, depending on the production rule \( p : A \rightarrow m \).

**First case:** \( m \) does not contain any non-terminal \( B =_1 A \). We let \( R_0(p) := m \), and \( R_1(p) \) and \( R_2(p) \) be the empty words.

**Second case:** \( m \) contains a unique non-terminal \( B \) with \( B =_1 A \) and \( m = m' B m'' \). We let \( R_1(p) := m' \) and \( R_2(p) := m'' \). Since we assume that \( A \neq_2 A \), the word \( m \) cannot contain two occurrences of non-terminals \( =_1 A \). In this case, \( R_0(p) \) is the empty word.

\[ \blacktriangleright \textbf{Claim 27.} \quad \text{For every } A \text{ such that } A \neq_2 A, \text{ we have:} \]

\[ L(G, A)^\downarrow = (\cup \alpha(L(G, R_1(p))))^\downarrow (\cup L(G, R_0(p))^\downarrow) (\cup \alpha(L(G, R_2(p))))^\downarrow \]

where the union extend to all production rules \( p \) with lefthand side \( B \) such that \( B =_1 A \).

Since the words \( R_0(p), R_1(p) \) and \( R_2(p) \) contain only non-terminals \( C \) with \( C <_1 A \) and \( C \neq A \), we have achieved our goal. By this we end the brief recall of the the approach described in [16] to compute downward closure.

Based on the above facts, we can construct intermediate CFG \( G' = (N', T', R', S') \), which helps to construct PDA for \( \mathcal{L}(G)^\downarrow \). For each non-terminal \( A \in N \), we introduce five non-terminals \( A_{\downarrow}, A_1, A_m, A_r \) and \( A_{\text{alph}} \) as well as dummy terminals \( a_l, a_r \) and \( a_{\text{alph}} \) in \( G' \).
Construction of $G'$ is shown in Algorithm 1 in Appendix D.3. The non-terminal $A_{\text{alph}}$ as well as the terminal $a_{\text{alph}}$ are used to simulate $\alpha(L(G, A))$. Likewise, the non-terminal $A_1$ ($A_r$) and the terminal $a_1$ ($a_r$) are used to simulate $\alpha(L(G, m')) (\alpha(L(G, m''))) when $m = m' \Delta m''$ in the production $A \rightarrow m$ and $C = 1$. The non-terminal $A_m$ is used when $m = A_1A_2$ for non-terminals $A_1 \neq 1, A_2 \neq 1$. Finally, $A_1$ works in place of $A$, and depending on $m$ in the production $A \rightarrow m$, gets rewritten either as $A_{\text{alph}}$ or $A_1A_mA_r$ or $A_1 \downarrow A_2 \downarrow$.

Constructing the PDA $A_{G,l}$ from $G, G'$. The PDA $A_{G,l} = (Q, T, \Sigma, \delta, \{q\}, S \downarrow, \{q\})$ from $G'$ and $G$ is constructed below. This works in the standard way of converting CFGs to PDA, keeping in mind the following. When $a_i$ (resp. $a_r$, $a_{\text{alph}}$) is the top of the stack, the PDA has a loop over symbols from $\alpha_i(A)$ (resp. $\alpha_r(A)$, $\alpha(L(G, A))$), in whichever state it is, at that time. $\alpha_i(A)$ (resp. $\alpha_r(A)$) is a set of symbols appearing in $\alpha(L(G, m')) (\alpha(L(G, m''))$ for all rules $C \rightarrow m'Bm''$ where $C = 1, A = 1, B$. The PDA stays in state $q$ when the top of stack contains a non-terminal of $A$. This non-terminal is popped and the right hand side $m$ of its production rule $A \rightarrow m$ is pushed on the stack. If the top of the stack is $i$) a terminal from $T$ which also happens to be the next input symbol, then we just pop it ii) a terminal of the form $a_{\text{alph}}, a_i$ or $a_r$, we move to state $q_{A_{\text{alph}}}, q_{A_i}$ or $q_{A_r}$ respectively and loop over $\alpha(L(G, A)), \alpha_i(A)$ or $\alpha_r(A)$ respectively and come back to $q$ non-deterministically. The PDA accepts in $q$ when the stack becomes empty. $A_1, A_r$ and $A_{\text{alph}}$ do not produce any further non-terminals while $A_m$ produces non-terminals $B$ where $B < 1$ and $B \neq 1$. Thus, the stack size is $\leq O(|N|)$.

\[D.3\quad \text{PSPACE-hardness in Theorem 14}\]

Our construction is inspired from \[27\] Lemma 8.6.

Let $M = (Q, \Sigma, \Gamma, \delta, q_0, B, F)$ be a linear bounded TM (LBTM). $\Sigma$ is the input alphabet, $\Gamma$ is the tape alphabet (contains $\Sigma$ and $B$ and some extra symbols) and $B$ stands for the blank symbol. $M$ uses at most $p(n)$ space for the input of size $n$, where $p(n)$ is a polynomial in $n$. The maximum number of distinct configurations of $M$ can be $\max(|Q|, |\Gamma|^{|w|+1})$, say $k_{\text{max}}$, assuming a configuration is represented as $w_1q_{w_2}$ where $w_1w_2$ is a word over $\Gamma$ that covers the entire tape (thus, we also include the spaces occupied by the blank symbols).

Consider constants $k_{\text{max}} = 2^{[\log k_{\text{max}}]}$ and $c_{\text{max}} = \log k_{\text{max}}$ to give precise bounds. Without loss of generality we assume that all halting computations of $M$ on $w$ take an even number of steps.

Given the LBTM $M$, we give the construction of PDAs $P_1$ and $P_2$ as follows. The input alphabet of $P_1, P_2$ is $\Gamma \cup Q \cup \{\#\,\}$, where $\#, \$ are new symbols. The stack alphabet for $P_1, P_2$ contain these symbols as well as symbols 0, 1, $\#$, 0, 1 are used to encode binary numbers to keep track of the number of configurations seen so far on any input. The idea is to restrict this number upto $k_{\text{max}}$.

$P_1$ and $P_2$ both generate words of the form $c_1\#c_2^R\#c_3^R\#\cdots\#c_k^R\#^s\$ where number of $\#$ is $k_{\text{max}}$, satisfying the following conditions.

1. Each $c_i$ is a valid configuration of $P$ and of size exactly $c_{\text{max}}$.
2. In $P_2$, $c_1$ is an initial configuration of $M$ on $w$ i.e. $q_0wB^\text{max}-(|w|+1)$.
3. In $P_1$, the last configuration before $\$ is an accepting configuration.
4. $c_i$ is a valid successor configuration of $c_i$ in $P_1$ for all odd $i$.
5. $c_i$ is a valid successor configuration of $c_i$ in $P_2$ for all even $i$.

This way, correctness of consecutive odd and even configurations are guaranteed by $P_2$ and $P_1$ respectively and $\mathcal{L}(P_1) \cap \mathcal{L}(P_2)$ is the set of accepting computations of $M$ on $w$ padded with copies of word $(\$^{\text{max}}\#)$ at the right end. The length of the words accepted by $P_1$ and
$P_2$ is $k_{\text{max}}$ (number of #s) + $c_{\text{max}} \cdot k_{\text{max}}$ (the size of each $c_i$ or $c_i^R$ is $c_{\text{max}}$, the padding with $\$\$ also has length $c_{\text{max}}$. There are $k_{\text{max}}$ of these $c_i$ or padded $\$\$.)

**Construction of $P_1$.**

(a) Initially we push $\log k_{\text{max}}$ zeros (where $k_{\text{max}}$ is the maximum possible number of configurations) and a unique symbol #$_1$ in the stack to keep track of number of #s read so far. #$_1$ is used as a separator in the stack between the counter values (encoded using 0,1) and the encoded configurations over $\Gamma, Q$. Thus, we initialize the value of the binary counter to $\log k_{\text{max}}$ zeros, with the separator #$_1$ on top.

(b) Let $L$ be $\{c_1 \# c_i^R \# | c_2$ is the successor configuration of $c_1$ in $M\}$. To realize $L$, we do the following. We read $c_1$ and check whether it is of the form $w_1qw'_1$ where $w_1, w'_1 \in \Gamma^*$ and $q \in Q$. While reading $c_1$ up to $q$, we push $w_1$ on its stack. As soon as we find $q \in Q$ in $c_1$, we store $q$ in the finite control state and read the next input symbol, say $X$.

- If $X$ is #, then there is no successor configuration, and we reject the word denoting that we have reached the end of the tape.
- If $\delta(q, X) = (p, Y, R)$, then we push $Yp$ onto the stack.
- If $\delta(q, X) = (p, Y, L)$, let $Z$ be on top of stack, then we replace $Z$ by $pZY$.

Again while reading $w'_1$, we push the symbols read onto the stack. Now the stack content will be (from bottom to top) the encoding of the counter, #$_1$, and $c_2$. After reading the # after $c_1$, we compare each input symbol with the top stack symbol (now we are reading $c_i^R$ in the input). If they differ, there is no next move. If they are equal, we pop the top stack symbol. When the top of stack is #$_1$, we read # after $c_i^R$ and accept. While reading $c_i$ (or $c_i^R$) we must also verify that the length of $c_i$ is exactly $c_{\text{max}}$. We can do this by adding polynomial size counter in the finite control state.

(c) We also need to increment the counter by 2 for the number of # (configurations) seen so far. We have read two configurations now in the above step. This can be done by popping the stack content starting from #$_1$ until we find 0. During this popping, we store in the finite control, the information of how many 1s we have popped. As soon as we find 0 on the top of the stack, replace it with 1 and push back as many number of 1s as we popped previously (this info is stored in the finite control), followed by the separator #$_1$. We need to do this step again to increment the counter since we saw two configurations (the two #s represent this) in $L$.

To construct the PDA $P_1$, we follow steps of [(a)] i.e., initializing the counter. Next it iterates over steps of [(b)] followed by [(c)], thus reading two consecutive $c_i \# c_i^R \#$ and incrementing the counter by 2 until we find the configuration of the form $\Gamma^* F \Gamma^*$ or the counter becomes full. If we stop the iteration due to accepting configuration and counter is not full yet (counter is full when we have all 1s in the stack), we pad $\$$^\text{\text{max}} \#$ until it becomes full. Otherwise if we stop the iteration due to the counter reaching $k_{\text{max}}$ value before getting accepting configuration, we stop and reject this sequence of configurations.

**Construction of $P_2$.** Similarly, we can construct PDA $P_2$: The difference here is that it starts with initial configuration i.e. $q_0wB^{c_{\text{max}}-|w|}$ followed by # (here also we increment the counter by 1), then it will iterate over $\{c_i^L \# c_2 \# | c_2$ is the successor configuration of $c_1$ in $M\}$ and increment the counter by 2 after each iteration. Then it will read accepting configuration non-deterministically (of course, followed by # and increment in the binary counter) and pad the string with $\$$^\text{\text{max}} \#$ until counter becomes full.

By padding the words accepted with extra $\$$^\text{\text{max}} \#s$ we are enforcing that $P_1$ and $P_2$ accepts words of same length. Having the blank symbols $B$ as part of the LBTM configuration is also motivated by this. Indeed, $w \in L(P_1) \cap L(P_2)$ iff the LBTM $M$ accepts $w$. Since the words accepted by $P_1, P_2$ have same constant length, $L(P_1) \cap L(P_2) \neq \emptyset$ iff $L(P_1)^\uparrow \cap L(P_2)^\downarrow \neq \emptyset$. 


On the Separability Problem of String Constraints

D.4 Proof of Claim 26

Suppose \( w \in L(G,A)^\downarrow \), then \( w \) is labelled by symbols from \( \alpha(L(G,A)) \). Clearly \( w \in (\alpha(L(G,A)))^\ast \).

Conversely, assume \( w \in (\alpha(L(G,A)))^\ast \). We prove by induction on the length of \( w \) that \( w \in L(G,A)^\downarrow \). If the length of \( w \) is 1, then trivially one can generate any word in \( L(G,A) \), which contains \( w \) as a subword. Assume the length of \( w \) to be \( k + 1 \). By induction hypothesis, there exists a word \( w' \in L(G,A) \) such that \( w[:k] \leq w' \) and \( w' \) is generated from \( A \). We construct \( w'' \in L(G,A) \) such that \( w \preceq w'' \). The derivation of \( w'' \) from \( A \) proceeds as follows. Start with the rule \( A \Rightarrow^+ mAm'Am'' \), and substitute the first \( A \) with \( w' \). For the second \( A \), substitute any word \( w''' \) s.t. \( A \Rightarrow_G^+ w''\). Then we obtain \( w \preceq w'' \), and hence \( w \in L(G,A)^\downarrow \).

D.5 Construction of \( G' \) and \( A_{G_\downarrow} \)

The PDA \( A_{G_\downarrow} \) from \( G, G' \). Now we construct the PDA \( A_{G_\downarrow} = (Q,T,N',\delta,q,\{q\}) \) from \( G' \) and \( G \). This works in the standard way of converting CFGs to PDA, keeping in mind the following. When \( a_l \) (resp. \( a_r, a_{\text{alph}} \)) is the top of the stack, the PDA has a loop over symbols from \( \alpha_l(A) \) (resp. \( \alpha_r(A), \alpha(L(G,A)) \)), in whichever state it is, at that time. The PDA stays in state \( q \) when the top of stack contains a non-terminal \( A \). This non-terminal is popped the right hand side \( m \) of its production rule \( A \rightarrow m \) is pushed on the stack. If the top of the stack is i) a terminal from \( T \) which also happens to be the next input symbol, then we just pop it ii) a terminal of the form \( a_{\text{alph}}, a_l \) or \( a_r \), we move to state \( q_{a_{\text{alph}}}, q_{a_l} \) or
We can construct a PDA of $L$ form $R$.

Proof. Let $G$ be a piece language over $\Sigma$ containing all words having exactly $(n-1)$ # symbols. We first show that $L \cap R$ can be written as a finite Boolean combination, except negation, of languages of the form $L_1\#L_2\#\ldots\#L_n$ where each $L_i$ is a piece language over $\Sigma$.

We prove inductively the representation of $L \cap R$ in terms of positive Boolean combinations of $n$ piece languages over $\Sigma$ having # in between them. As a base case, let $L$ be a piece language over $\Sigma$ containing exactly $n$ # symbols. Then, $L \cap R$ is a finite union of languages of the form $L_i\#L_j\#\ldots\#L_n$ where each $L_i$ is a piece language over $\Sigma$.

$q_{A_n}$ respectively and loop over $\alpha(L(G,A))$, $\alpha_l(A)$ or $\alpha_r(A)$ respectively and come back to $q$ non-deterministically. The PDA accepts in $q$ when the stack becomes empty.

Complexity. We analyze the complexity of constructing $G'$ as well as $A_{G'}$.

- Construction of $G'$:
  - for every $A, B \in N$, checking $A \prec B$ reduces to checking membership of $a_1a_2$ in $L(G_1, B)$ where $G_1$ is obtained from $G$ by replacing productions $C \to a$ of $G$ with $C \to \epsilon, C \to a_1 | a_2$. $G_1$ is constructed in polynomial time from $G$. Similarly, we can check if $A \prec B$ and $a \in \alpha(L(G, A))$.
  - for every $A \in N$, five non-terminals and three terminals are introduced. Hence, $G'$ has $5|N|$ non-terminals and $|T| + 3|N|$ terminals, which is again polynomial in the input $G'$.
  - each step of the algorithm [1] can be computed in at most polynomial time and the algorithm has loops whose sizes range over terminals, non-terminals and production rules of $G$.

- PDA $A_{G'}$: Each non-terminal $A$ has corresponding non-terminals $A_l, A_r, A_{\text{alph}}$ or $A_m$ in $G'$. Out of these, $A_l, A_r$ and $A_{\text{alph}}$ do not produce any further non-terminals while $A_m$ produces non-terminals $B$ where $B < A$ and $B \neq A$. Thus, the stack size is bounded above by $O(|N|)$.

D.6 Proof of Lemma 24

Let $G = (N, T, R, S)$, we construct CFG $G'$ which accepts subset of $L(G)$ and includes $\min(L(G))$. Clearly $L(G) \upharpoonright = L(G') \upharpoonright$. Let $G'$ be $(N \times \{0, \ldots, |N|\}, T, R', (S, |N|))$ where production rules are defined as following:

- $(A, 0) \to a$ if $A \to a$
- $(A, i) \to (B, i - 1)(C, i - 1)$ for all $i \in [1, |N|]$ if $A \to BC$
- $(A, i) \to (A, i - 1)$ for all $i \in [1, |N|]

We can construct a PDA $P$ for $L(G')$ in polynomial time. It is easy to see that the PDA uses at most $|N| + 1$ height of the stack. It accepts all words of $L(G)$ having derivation with at most $|N| + 1$ steps. We claim that $L(G')$ contains $\min(L(G))$. If not, then there exists $w \in \min(L(G))$ with at least $|N| + 2$ derivation steps i.e. some non-terminal $A$ appears at least two times in derivation tree say at level $i$ and $j > i$. If we replace the subtree rooted at level $j$ by subtree at level $i$, we get a proper subword of $w$ (wlog we assume $G$ is in CNF and it does not have useless symbols and $\epsilon$–productions), which is also in $L(G)$, contradicting our assumption that $w \in \min(L(G))$. Finally we can obtain a PDA for $L(G') \upharpoonright$ by adding self loops on each state of $P$ for each symbol $a \in \Sigma$ as $L(G) \upharpoonright = L(G') \upharpoonright$.

D.7 Proof of Lemma 15

Given two SL formulas $\Psi$ and $\Psi'$, with $x_1, \ldots, x_n$ as their set of variables. Let $A$ and $A'$ be two OMPas such that $L(A) = \text{Encode}(\Psi)$ and $L(A') = \text{Encode}(\Psi')$. Then, $\Psi$ and $\Psi'$ are separable by an $n$-PosPTL iff $A$ and $A'$ are separable by a PosPTL.

Proof. Let $L(A)$ and $L(A')$ be separable by a PosPTL $L$ over $\Sigma \cup \{\#\}$. Consider a regular language $R \subseteq (\Sigma \cup \{\#\})^*$ containing all words having exactly $(n-1)$ # symbols. We first show that $L \cap R$ can be written as a finite Boolean combination, except negation, of languages of the form $L_1\#L_2\#\ldots\#L_n$ where each $L_i$ is a piece language over $\Sigma$.

We prove inductively the representation of $L \cap R$ in terms of positive Boolean combinations of $n$ piece languages over $\Sigma$ having # in between them. As a base case, let $L$ be a piece language over $\Sigma$ containing exactly $n$ # symbols. Then, $L \cap R$ is a finite union of languages of the form $L_i\#L_j\#\ldots\#L_n$ where each $L_i$ is a piece language over $\Sigma$.
language $\Sigma^*a_1\Sigma^* \ldots \Sigma^*a_k\Sigma^*$. Let $S$ be a finite set containing only the words of the form $w$ such that $a_1a_2 \ldots a_k \leq w$ and the symbol $\#$ appears exactly $(n-1)$-times in $w$. Then $L \cap R$ can be written as $\bigcup_{w \in S} \Sigma^*b_1\Sigma^*b_2 \ldots \Sigma^*b_k\Sigma^*$, where $w = b_1 \ldots b_k$, which is of the form $\bigcup_{i=1}^{|S|} L_{1i} \# L_{12} \# \ldots \# L_{1n}$ where $L_{1i}$ is a piece language over $\Sigma$.

Now assume that $L$ is the form of the $L_1 \cap L_2$ (resp. $L_1 \cup L_2$). It is easy to see that $L \cap R$ is equivalent to $(L_1 \cap R) \cap (L_2 \cap R)$ (resp. $(L_1 \cap R) \cup (L_2 \cap R)$). Thus we can use our induction hypothesis to show that $L \cap R$ is indeed a Boolean combination (using union and intersection) of languages of the form $L_1 \# L_2 \# \ldots \# L_n$ where $L_i$ is a piece language over $\Sigma$.

Now, we are ready to construct an $n$-PosPTL that separates $L(\Psi)$ and $L(\Psi')$. First of all, observe that $L(A) \subseteq L \cap R$ and $L(A') \cap (L \cap R) = \emptyset$ (from the definition of $A$ and $A'$). Since $L \cap R$ can be written as a positive Boolean combination of languages of the form $L_1 \# L_2 \# \ldots \# L_n$, where $L_i$'s are piece languages over $\Sigma$, we can easily construct an $n$-PosPTL over $\Sigma$ that separates the two formulas $\Psi$ and $\Psi'$ as a finite positive Boolean combination of the product of these $L_i$'s. This is similar to Lemma 10 where we used the fact that a finite Boolean combination of the product of piece languages is a PTL over an $n$-tuple alphabet. The proof of the other direction follows similarly. ▶

D.8 Proof of claim in section 5.2, line 421

Converting SL with functional transducers to right-sided SL

Lemma 28. Let $\Psi$ be an SL formula where all relational constraints have only functional transducers. Then we can obtain a right-sided SL formula $\Psi'$ such that $\Psi'$ is satisfiable iff $\Psi$ is.

Proof. Assume that the SL formula $\Psi$ with functional transducers is given as

$$\Psi = \bigwedge_{i=1}^n x_i \in L(A_i) \land \bigwedge_{i=1}^k \varphi_i$$

such that $\varphi_i$ is of the form $T_i(t_i) = x_i$, and $T_i$ defines a (partial) function. Without loss of generality, assume that each $t_i$ is a sequence of variables. By the SL condition, if a variable $x_j$ appears in $t_i$ then $j > i$. Notice that $\Psi$ need not be rightsided, since we can have $\varphi_k$ as the constraint $x_h = T_h(t_h)$ with $t_h = x_j$, $j > h$ and also $\varphi_j$ being $x_j = T_j(t_j)$ violating the condition that none of the output variables $x_1, \ldots, x_k$ can be present in any of the inputs $t_1, \ldots, t_k$. The variables $x_{k+1}, \ldots, x_n$ are independent and do not appear in the left side (or outputs) of any of the transductions.

We first explain the idea of the proof. We construct a formula $\Psi'$ in which the variables appearing in the inputs terms $t_1', \ldots, t_k'$ of all the transductions in $\Psi'$ are independent variables of $\Psi'$. Thus, each of $t_1', \ldots, t_k'$ will be sequences of variables from $x_{k+1}, \ldots, x_n$. The idea is to apply substitution inductively starting from $\varphi_k$ to $\varphi_1$, obtaining in $\Psi'$, the constraints $\varphi_k', \ldots, \varphi_1'$ s.t. $\Psi'$ are rightsided. Note that by the SL definition, $t_k$ can have only variables from $x_{k+1}, \ldots, x_n$. Now consider $\varphi_{k-1}$ given by $x_{k-1} = T_{k-1}(t_{k-1})$. We know that $t_{k-1}$ can have only variables from $x_k, x_{k+1}, \ldots, x_n$. We can replace all occurrences of $x_k$ in $t_{k-1}$ with $T_k(t_k)$ obtaining $x_{k-1} = T_{k-1}(t_{k-1}[x_k \mapsto T_k(t_k)])$. Likewise, we can replace occurrences of $x_{k-1}, x_k$ in $t_{k-2}$ respectively with $T_{k-1}(t_{k-1})$ and $T_k(t_k)$. We proceed iteratively for $\varphi_{k-3}$ to $\varphi_1$. 
We give the formal construction by induction on \( n \). For the base case, \( \varphi_k \) is same as \( \varphi_k \) since \( t_k \) is over the independent variables \( \{x_{k+1}, \ldots, x_n\} \). The same holds for all \( \varphi_k \) s.t. the input \( t_h \) is over \( \{x_{k+1}, \ldots, x_n\} \).

For the inductive step, let \( j \) be the largest index s.t. the input \( t_j \) in \( \varphi_j \) contains dependent variables (from \( \{x_{j+1}, \ldots, x_k\} \)). Let \( \varphi_j \) be given by \( x_j = T_j(t_j) \) where \( t_j = t_j[1] \ldots t_j[m] \), where \( t_j[1], \ldots, t_j[m] \in \{x_{j+1}, \ldots, x_k, x_{k+1}, \ldots, x_n\} \). Consider an accepting run in \( T_j \) visiting states \( q_0, \ldots, q_m \) s.t. the part of the run between \( q_{i-1} \) and \( q_i \) processes the input \( t_j[i] \) if there are multiple accepting runs for the same input word in \( T_j \), the sequence \( s \) of states \( q_0, \ldots, q_m \) can differ, but the output will be the same since \( T_j \) is functional. Let \( T_j(q_0, q_1), \ldots, T_j(q_{m-1}, q_m) \) represent transducers induced from \( T_j \), s.t. \( T_j(q_0, q_1) \) reads the first variable \( t_j[1] \) and produces output (say \( o_1 \)), \( T_j(q_1, q_2) \) reads the second variable \( t_j[2] \) and produces output (say \( o_2 \)), and so on till \( T_j(q_{m-1}, q_m) \) reads the \( m \)th variable \( t_j[m] \) and produces output (say \( o_m \)). Let \( T_j^{s,i} \) denote the induced transducer \( T_j(q_{i-1}, q_i) \) for \( 1 \leq i \leq m \) (i.e., the transducer \( T_j \) with \( q_{i-1} \) as initial state and \( q_i \) as a final one). Then, for a fixed sequence of states \( q_0, \ldots, q_m \), we can write \( x_j \) as \( T_j^{s,1}(t_j[1])T_j^{s,2}(t_j[2]) \ldots T_j^{s,m}(t_j[m]) \).

Note that the concatenation of transducers \( T_j^{s,1}, \ldots, T_j^{s,m} \) over inputs \( t_j[1], \ldots, t_j[m] \) can be replaced by a single transducer \( T_j' \) over the input \( t_j = t_j[1] \ldots t_j[m] \) producing output \( o_1 \ldots o_m \), since transducers are closed under concatenation.

Assume that \( t_j[\ell] = x_p \) is a dependent variable, \( p \in \{j + 1, \ldots, k\} \). Then we can replace \( T_j^{s,\ell}(t[\ell]) \) with \( T_j^{s,\ell}([t_p]) \). Now, we apply such substitution to all the variables appearing in \( t \). This will result in that \( t' \) is a sequence over independent variables \( \{x_{k+1}, \ldots, x_n\} \).

Thus, we can rewrite \( x_j \) as \( T_j^{s,1}(t_{p_1}(t_{p_1}))T_j^{s,2}(t_{p_2}(t_{p_2})) \ldots T_j^{s,m}(t_{p_m}(t_{p_m})) \) where \( p_1, \ldots, p_m \) is defined such that \( t_j[\ell] = x_{p_\ell} \) for all \( \ell \in \{1, \ldots, m\} \), and \( T_{p_\ell} \) is the transducer given by \( T_{p_\ell} \) in case \( p_\ell \in \{j + 1, \ldots, k\} \), and is the identity otherwise. Functional transducers are closed under composition, hence we can obtain a functional transducer \( T_j'' \) equivalent to each of the compositions \( T_j^{s,1} \circ T_j^{s,2} \cdots T_j^{s,m} \) for \( 1 \leq i \leq m \). This expression for \( x_j \) is for a fixed accepting path through states \( q_0, \ldots, q_m \). Considering the general case of any accepting path over an accepting sequence \( q_{i_0} \ldots q_{i_m} \), we can write \( \varphi_j'' \) as

\[
\varphi_j'' = \bigvee_{s = q_{i_0} \ldots q_{i_m}} [x_j = T_j^{s,1}(t_{p_1}(t_{p_1}))T_j^{s,2}(t_{p_2}(t_{p_2})) \ldots T_j^{s,m}(t_{p_m}(t_{p_m}))]
\]

Notice again that \( t_{p_1}, t_{p_2}, \ldots, t_{p_m} \) is a sequence over independent variables, and the union and concatenation above can be replaced by a single transducer \( T_j' \). \( T_j' \) is the transducer obtained by the concatenation of the following transducers \( (T_j^{s,1}(t_{p_1})) \cdot (\epsilon, \$(}) \cdot (T_j^{s,2}(t_{p_2})) \cdot (\epsilon, \$(}) \cdots (\epsilon, \$(}) \cdot (T_j^{s,m}(t_{p_m})) \) where \( \$ \) is a special symbol not in \( \Sigma \). Now let \( T_j'' \) be the transducer obtained by taking the union of \( T_j' \) where \( s \) is a sequence of states of \( T_j \) of length \( m \). Now we can rewrite \( \varphi_j' \) as \( \varphi_j''' \) defined by

\[
x_j = T_j'(t_{p_1} \cdot t_{p_2} \cdot \ldots \cdot t_{p_m})
\]

We can now replace the occurrence of \( \$ \) by a fresh variable \( z \) not appearing in \( \Psi \) to obtain the constraint \( \varphi_j' \) defined by

\[
x_j = T_j'(t_{p_1} \cdot z \cdot t_{p_2} \cdot z \ldots \cdot t_{p_m})
\]

We now proceed with the next largest \( j \) s.t. the input of \( \varphi_j \) is replaced by \( \varphi_j' \) and has now has a dependent variable. We repeat this procedure until we get rid of all the dependent variables as input of the relational constraints. Let \( \Psi'' \) the resulting formula.

\[
\Psi'' = \bigwedge_{i=1}^{n} x_i \in \mathcal{L}(A_i) \land \bigwedge_{i=1}^{k} \phi_i
\]
We can now define the formula $Ψ'$ as the conjunction of $Ψ''$ and $z \in \{\}$.

Based on the construction of $Ψ'$, we have the following claims.

- **Lemma 29.** $Ψ'$ is a rightsided SL formula.

  **Proof.** By construction, we have
  
  $$Ψ' = z \in \{\} \land \bigwedge_{i=1}^{n} x_i \in \mathcal{L}(A_i) \land \bigwedge_{i=1}^{k} \varphi_i'$$

  The variables $z, x_{k+1}, \ldots, x_n$ are clearly independent in $Ψ'$ as well. Consider a $\varphi_j'$.

  As described above, we know that the input of the transduction $\varphi_j'$ is a sequence over independent variables (thanks to composition of transducers). All input variables are among \{z, x_{k+1}, \ldots, x_n\}, the independent variables. Hence $Ψ'$ is rightsided.

- **Lemma 30.** The formula $Ψ'$ constructed is satisfiable iff $Ψ$ is.

  **Proof.** The equivalence of each $\varphi_j'$ with $\varphi_j$ ensures that we have the same functions computed by the transductions, and hence all the inputs and outputs agree in each of the old and new transductions. This, along with the fact that independent variables remain untouched (with the exception of the fresh variable $z$ which has as evaluation $\{\}$), ensure that we have the same valuation for $x_1, \ldots, x_n$ in both $Ψ$ and $Ψ'$.

Thus, we have shown that, given a SL string constraint having only functional transducers, we can compute an equisatisfiable rightsided SL string constraint.

### D.9 Details for Lemma [19]

Given a right-sided SL formula $Ψ$ over $Σ$, with $x_1, x_2, \ldots, x_n$ as its set of variables, we wish to construct, in polynomial time, a 2NFT $A_Ψ$ such that $R(A_Ψ)=\{(u_1\#u_2\#\cdots\#u_n, w_1\#w_2\#\cdots\#w_n)|u_1\#u_2\#\cdots\#u_n \in \text{Encode}(\mathcal{L}(Ψ))$ and $w_i=u_i$, if $x_i$ is an independent variable $\}$.

To recall, $Ψ$ is of the form $\bigwedge_{i=1}^{n} y_i \in \mathcal{L}(A_i) \land \bigwedge_{i=1}^{k} (y_i, t_i) \in \mathcal{R}(T_i)$ with $y_1, \ldots, y_n$ is a permutation of $x_1, \ldots, x_n$. $\pi : [1, n] \rightarrow [1, n]$ is the mapping that associates to each index $i \in [1, n]$, the index $j \in [1, n]$ s.t. $x_i = y_j$ (or $x_i = y_{\pi(i)}$).

As already mentioned above, given $n$ blocks separated by $\#$, $A_Ψ$ treats block $i$ as valuation of $x_i$ if $x_i$ is an independent variable, for all $i$. For other (dependent) variables, say $x_j$, $A_Ψ$, computes its valuation based on the valuation of variables given in input blocks. $A_Ψ$ considers the ordering of variables as $x_1, \ldots, x_n$ for input and output. Thus if the input is $u_1\#u_2\#\cdots\#u_n$, then $A_Ψ$ tries to find satisfying assignment $η$ for $Ψ$, and output for this input is given as $η(x_1)\#\cdots\#η(x_n)$ s.t. $η \models Ψ$ and $η(x_i) = u_i$ for all independent variables $i$.

**Working of $A_Ψ$:** It produces valuation of $x_1, x_2, \ldots, x_n$ in sequence. Depending on if $x_i$ is dependent or independent variable, process given in Lemma [19] is followed by $A_Ψ$. To execute the process, $A_Ψ$ uses following set of states and transitions.

**Set of states:**

- For each $x_i$, if $x_i$ is an independent variable, we have following set of states used to produce $x_i$ in the output:
\{ (q_0(x_i), r) \mid r \text{ is an initial state of } A_{π(i)} \} \\

The first set indicates that now we want to output value of \( x_i \) in the run, which is present in \( i^{th} \) block in the input. So that we first move to \( \uparrow \), and \( A_\Psi \) remains in state \( (q_0(x_i), r) \) while performing this operation, i.e. moving toward left end.

\{ (q_j(x_i), r) \mid 0 \leq j < i, r \text{ is an initial state of } A_{π(i)} \} \\

These states are visited after \( (q_r(x_i), r) \), specifically when \( A_\Psi \) is in \( (q_j(x_i), r) \), it represents that we have already scanned \( j \) blocks in the input, and currently reading \((j + 1)^{th}\) block.

\{ (q_r(x_i), r) \mid r \in Q_{A_{π(i)}} \} \\

With the help of above states, \( A_\Psi \) reaches the desired \( i^{th} \) block in the input and in the beginning of this block, state is \( (q_r(x_i), r) \), where \( r \) is initial state of \( A_{π(i)} \). Later \( A_\Psi \) simulates the transitions of \( A_{π(i)} \) to check the membership constraint \( y_{π(i)} \in A_{π(i)} \) in the second component of states of \( A_\Psi \) with the help of third set.

For each \( x_i \), if \( x_i \) is a dependent variable, we have following set of states used to produce \( x_i \) in the output:

\{ (q_r(x_i, x_{i_1}, j)) \mid t_{π(i)}[j] = x_{i_1}, 1 \leq j \leq t_{π(i)}, p \in Q_{T_{π(i)}}, r \in Q_{A_{π(i)}} \} \\

The first set of states are similar to above case. \( \uparrow \) present in state name indicates that \( A_\Psi \) is in process of moving left to reach \( \uparrow \) and the next, we want to read \( x_{i_1} \) from \( t_{π(i)} \), position \( j \) (i.e. \( t_{π(i)}[j] = x_{i_1} \)), to produce part of valuation of \( x_i \). The reading of value of \( x_{i_1} \) must start from states \( p \) in \( T_{π(i)} \), while its output must match with the word starting from \( r \) in \( A_{π(i)} \). Here the third parameter \( j \) is important since it remembers the progress of \( t_{π(i)} \), how many variables from it are already processed.

\{ (q_k(x_i, x_{i_1}, j), (p, r)) \mid 0 \leq k \leq i, t_{π(i)}[j] = x_{i_1}, p \in Q_{T_{π(i)}}, r \in Q_{A_{π(i)}} \} \\

The second set of states are visited to count the number of \#s before reading value of \( x_{i_1} \). Once this count (presented by a subscript of \( q \)) reaches \( i_j - 1 \), we are in block \( i_j \). Then \( A_\Psi \) simulates the transitions of \( T_{π(i)} \) starting from \( p \) and in parallel simulates the transitions of \( A_{π(i)} \) starting from state \( r \) on the output produced by \( T_{π(i)} \).

d) special state \( q^{final} \) to mark the end.

**Initial states:**

\{(q_0(x_1, x_{1_1}, 1), (p_0, r_0)) \mid x_1 \text{ is dependent }, t_{π(1)} = x_{1_1}, p_0, r_0 \text{ are initial states of } T_{π(1)}, A_{π(1)} \}

\cup \{(q_0(x_1), r_0) \mid x_1 \text{ is independent }, r_0 \text{ initial of } A_{π(1)} \}

**Final states:** \( \{q^{final}\} \)

**Set of transitions:** For all \( i \in [1, n] \):

1. \( (q_r(x_i), r) \xrightarrow{\#_{[\alpha], r-1}} (q_r(x_i), r) \)

This loop helps to reach \( \uparrow \) in the input.

2. Once we reach the left end \( \uparrow \) from \( (q_r(x_i), r) \):

\( (q_r(x_i), r) \xrightarrow{\#_{[\alpha], r+1}} (q_0(x_1), r) \)

3. To read \( x_i \) from input, we need to skip \( m \) many \#s using following transitions. For all \( 1 \leq j \leq i - 2 \):

\( (q_j(x_i), r) \xrightarrow{\#_{[\alpha], r+1}} (q_j(x_i), r) \text{ for any } \alpha \in \Sigma \)

\( (q_j(x_i), r) \xrightarrow{\#_{[\alpha], r+1}} (q_{j+1}(x_i), r) \)
On the Separability Problem of String Constraints

$(q_{i-1}(x_i),r) \xrightarrow{\alpha | r, 0} (q_i(x_i), r)$ for any $\alpha \in \Sigma \cup \{\#, -\}$

These transitions help in counting the number of blocks separated by $\#$. It stops when $A_\Psi$ reaches the $i$th block.

- Once $A_\Psi$ reaches the $i$th block, it needs to simulate the transitions of $A_{\pi(i)}$: If $x_i$ is independent variable, and current state is $r$ while checking membership in $A_{\pi(i)}$:

$(q_i(x_i), r) \xrightarrow{\alpha | r, 1} (q_i(x_i), r')$ if $(r, \alpha, r') \in \delta_{A_{\pi(i)}}$, for any $\alpha \in \Sigma$

Above transition simulates $A_{\pi(i)}$ in second component while producing the same output. After it finishes the block $x_i$, it moves ahead to produce $x_{i+1}$ next producing separator $\#$ in the output:

1. If $x_{i+1}$ is independent variable, and $r_0$ is initial state of $A_{\pi(i+1)}$, $r$ is final state of $A_{\pi(i)}$:

$(q_i(x_i), r) \xrightarrow{\# | r, -1} (q_i(x_{i+1}), r_0)$

2. If $x_{i+1}$ is dependent variable, and $r_0$ is initial state of $A_{\pi(i+1)}$, $p_0$ is initial state of $T_{\pi(i+1)}$, $t_{\pi(i)}[1] = x_{i+1}$, $r$ is final state of $A_{\pi(i)}$:

$(q_i(x_i), r) \xrightarrow{\# | r, -1} (q_i(x_{i+1}, x_{i+1}, 1), (p_0, r_0))$

- If $x_i$ is dependent variable, $r_0$ is initial state of $A_{\pi(i)}$, $p_0$ is initial state of $T_{\pi(i)}$, and $t_{\pi(i)}[1] = x_i$:

$(q_i(x_i, x_i, 1), (p_0, r_0)) \xrightarrow{\alpha | r, 1, -1} (q_i(x_i, x_i, 1), (p_0, r_0))$ for any $\alpha \in \Sigma$

This transition moves $A_\Psi$ towards $\vdash$. Once $\vdash$ is reached, it starts counting the number of blocks separated by $\#$s.

$(q_i(x_i, x_i, 1), (p_0, r_0)) \xrightarrow{\# | r, 1, -1} (q_0(x_i, x_i, 1), (p_0, r_0))$

- If $x_i$ is dependent variable, and $t_{\pi(i)}[j] = x_{ij}$, for any $0 \leq k \leq i_j - 2$: (Counting of $\#$ separated blocks)

$(q_k(x_i, x_i, j), (p, r)) \xrightarrow{\alpha | r, 1} (q_k(x_i, x_i, j), (p, r))$ for any $\alpha \in \Sigma$

$(q_k(x_i, x_i, j), (p, r)) \xrightarrow{\# | r, 1} (q_{k+1}(x_i, x_i, j), (p, r))$

- If $x_i$ is dependent variable, and $t_{\pi(i)}[j] = x_{ij}$, $k = i_j - 1$ (When the desired block is reached in the input):

$(q_k(x_i, x_i, j), (p, r)) \xrightarrow{\alpha | r, 0} (q_{k+1}(x_i, x_i, j), (p, r))$

- If $x_i$ is dependent variable, and $t_{\pi(i)}[j] = x_{ij}$, $k = i_j$ (Simulating transitions of $T_{\pi(i)}$ on input, while $A_{\pi(i)}$ on output):

1. if $(p, (b, a), p') \in \delta_{T_{\pi(i)}}$ and $(q, b, q') \in \delta_{A_{\pi(i)}}$, then

$(q_k(x_i, x_i, j), (p, r)) \xrightarrow{a | b, 1} (q_k(x_i, x_i, j), (p', r'))$

2. if $(p, (b, c), p') \in \delta_{T_{\pi(i)}}$ and $(q, b, q') \in \delta_{A_{\pi(i)}}$, then

$(q_k(x_i, x_i, j), (p, r)) \xrightarrow{a | b, 0} (q_k(x_i, x_i, j), (p', r'))$

3. if $|t_{\pi(i)}| > j$, $t_{\pi(i)}[j + 1] = x_{i_{j+1}}$ (If the current variable is not the last in the term, $A_\Psi$ proceeds with next variable in term ).
\[(q_k(x_i, x_{i+1}, j), (p, r)) \xrightarrow{\#|\epsilon, -1} (q_p(x_i, x_{i+1}, j + 1), (p, r))\]

4. if \(|t_\pi(i)| = j\), \(r\) and \(p\) are final states of \(A_{\pi(i)}\) and \(T_{\pi(i)}\) respectively (If the current variable is the last in the term, \(A_\phi\) has finished producing \(x_i\) and proceeds with next variable \(x_{i+1}\) after producing \# as separator),

a. if \(x_{i+1}\) is an independent variable, and \(r_0\) is initial state of \(A_{\pi(i+1)}\)
\[(q_k(x_i, x_{i+1}, j), (p, r)) \xrightarrow{\#|\epsilon, -1} (q_p(x_i+1), r_0)\]

b. if \(x_{i+1}\) is dependent variable, and \(r_0\) is initial state of \(A_{\pi(i+1)}\), \(p_0\) is initial state of \(T_{\pi(i+1)}\), \(t_\pi(i)[1] = x_{i+1}\)
\[(q_k(x_i, x_{i+1}, j), (p, r)) \xrightarrow{\#|\epsilon, -1} (q_p(x_i+1, x_{i+1}, 1), (p_0, r_0))\]

- If \(x_n\) is independent variable and \(r\) is a final state of \(A_{\pi(n)}\)
\[(q_n(x_n), r') \xrightarrow{\#|\epsilon, +1} q_{final}\]

- If \(x_n\) is dependent variable, and \(t_\pi(n)[j] = x_{n_j}\), \(k = n_j\), \(|t_\pi(n)| = j\), \(r\) and \(p\) are final states of \(A_{\pi(n)}\) and \(T_\pi(n)\) respectively,
\[(q_k(x_n, x_{n_j}, j), (p, r)) \xrightarrow{\#|\epsilon, +1} q_{final}\]

- finally, since the acceptance condition is to reach final state after reading \(\vdash\)
\[q_{final} \xrightarrow{\vdash|\epsilon, +1} q_{final}\]