Semi-linear VASR for Over-Approximate Semi-linear Transition System Reachability

Nikhil Pimpalkhare(✉) and Zachary Kincaid

Princeton University, Princeton, NJ 08544, USA
{np6641,zkincaid}@cs.princeton.edu

Abstract. This paper introduces Semi-Linear Integer Vector Addition Systems with Resets (SVASR). A SVASR is a labeled transition system in which the states are finite-dimensional integer-valued vectors and which transitions from one state to another by applying an orthogonal projection followed by a translation drawn from a semi-linear set. We give a polynomial-time reduction of SVASR reachability to that of Integer Vector Addition Systems with Resets.

We then consider the use of SVASRs for over-approximating the reachability relation of transition systems in which the transition relation is a semi-linear set. We show that any semi-linear transition system has a "best" SVASR that simulates its behavior, called its SVASR-reflection. The dimension of the SVASR-reflection of a semi-linear transition system T with states is exponential in the number of states; however, we show that the over-approximate reachability induced by T's SVASR-reflection can be computed in polynomial time.

Keywords: Vector Addition Systems · Linear Integer Arithmetic

1 Introduction

Vector Addition Systems (VAS) are a widely-studied class of infinite-state transition systems. Classically, states of these systems are finite vectors over the naturals and transitions increment the state by a translation vector over the integers drawn from a finite set. Haase and Halfon initiated the study of *integer* VAS, in which states are integer-valued vectors, and showed that the reachability relation for integer VAS is definable in linear integer arithmetic even in the presence of states and resets [3]. In a separate line of work, Piskac and Kunčak showed that linear integer arithmetic is effectively closed under star—i.e., if $F(x)$ is a linear integer arithmetic formula, then the set $\{\sum_{i=1}^{n} m_i : n \in \mathbb{N}, \forall i.m_i \models F\}$ is LIA-definable [9].

In this paper, we study a common generalization of these two lines of work: integer vector addition systems with resets in which the set of translation vectors is infinite but LIA-definable, or equivalently a semi-linear set. We refer to such

L. Kovács and A. Sokolova (Eds.): RP 2024, LNCS 15050, pp. 154–166, 2024.
https://doi.org/10.1007/978-3-031-72621-7_11

transition systems as Semi-Linear Integer VAS with Resets (SVASR). We show that that reachability of SVASR reduces to that of integer VAS with resets (VASR).

Next, we consider the application of SVASRs to over-approximate the reachability relation of semi-linear transition systems—transition systems for which the transition relation is a semi-linear set (or equivalently LIA-definable). The reachability problem for this class of transition system is undecidable, since it generalizes counter machines. Following the strategy of [8,10], we can over-approximate the reachability relation of a semi-linear transition system by (1) computing a SVASR that simulates it and (2) computing the inverse image of the reachability relation of this SVASR under the simulation. We show that every semi-linear transition system has a best abstraction as a SVASR called its *SVASR reflection*. It is best in the sense that the reflection's transition relation over-approximates the semi-linear transition system's at least as precisely as any other SVASR.

The dimension of the SVASR-reflection of a semi-linear transition system is exponential in the size of its alphabet, so a direct attempt to compute over-approximate reachability via the SVASR-reflection results in an exponential-space algorithm. We show that we can compute an equivalent formula in polynomial time by avoiding explicit computation of the SVASR reflection.

In summary, we introduce a new extension of vector addition systems, the SVASR, and show that its reachability reduces to that of VASR. We then propose a practical technique to computing over-approximate reachability of semi-linear transition systems using its SVASR reflection; the salient feature of this technique is that the computed over-approximation is guaranteed to be at least as precise as that induced by any other SVASR abstraction.

2 Background

A semi-linear set [7] in a $\mathbb{Z}$-module V is the finite union of linear sets in V. A linear set S in V is generated by a *base point* b in V and a sequence of *periods* $p_1, \ldots, p_n$ in V:

$$S \triangleq \{b + \lambda_1 p_1 + \cdots + \lambda_n p_n : \lambda_1, \ldots, \lambda_n \in \mathbb{N}\}$$

The generator representation of a semi-linear set is not unique. For convenience, we assume that every semi-linear set has a canonical generator which we refer to as a **basis**. A basis over vector space V is an element of $(V \times (V)^*)^*$. Each element of this representation is the generator representation of a linear set; that is, a base pointer followed by a sequence of periods. Let $B(S)$ refer to the canonical generator representation of semilinear set S and let $S(B)$ be the semilinear set defined by basis B.

A **labeled** transition system[1] T over a set of variables X and alphabet Σ is a pair $\langle \mathbb{Z}^X, \rightarrow_T \rangle$ where $\mathbb{Z}^X$ is a state space and $\rightarrow_T \subseteq \mathbb{Z}^X \times \Sigma \times \mathbb{Z}^X$ is a labeled transition relation. We use the following notation:

[1] We restrict our attention to transition systems in which the state space is a finite-dimensional module over the integers.

- For a character $s \in \Sigma$, write $\rho \xrightarrow{s}_T \rho'$ if $\langle \rho, s, \rho' \rangle$ belongs to $\rightarrow_T$
- For a word $s_1 \dots s_n \in \Sigma^*$, write $\rho \xrightarrow{s_1 \dots s_n}_T \rho'$ if there exists a sequence of states $\rho_0 \dots \rho_n$ such that $\rho = \rho_0 \xrightarrow{s_1}_T \dots \xrightarrow{s_n}_T \rho_n = \rho'$
- For a language $L \subseteq \Sigma^*$, write $\rho \xrightarrow{L}_T \rho'$ if $\rho \xrightarrow{w}_T \rho'$ for some word $w \in L$

Various control features can be encoded using reachability constrained to a given language of paths. For instance, the reachability relation of a vector addition system with states (or with a pushdown stack) corresponds to the reachability relation of a vector addition system constrained to a regular language (or context-free language).

A linear simulation between labeled transition systems $T = \langle \mathbb{Z}^X, \rightarrow_T \rangle$ and $U = \langle \mathbb{Z}^Y, \rightarrow_U \rangle$ over the same alphabet Σ is a linear function $f : \mathbb{Z}^X \rightarrow \mathbb{Z}^Y$ such that for any $\rho, \rho' \in \mathbb{Z}^X$, if $\rho \xrightarrow{s}_T \rho'$ then $f(\rho) \xrightarrow{s}_U f(\rho')$. For any language L, if we can compute the L-reachability relation $\xrightarrow{L}_U$ of U, then we can over-approximate the L-reachability relation of T as $\left\{ \langle \rho, \rho' \rangle : f(\rho) \xrightarrow{L}_U f(\rho') \right\}$ (which must contain $\xrightarrow{L}_T$).

A labeled integer vector addition system with resets (**VASR**) over variables X and alphabet Σ is a labeled transition system $\mathcal{V} = \langle \mathbb{Z}^X, \rightarrow_{\mathcal{V}} \rangle$ such that for each character $s \in \Sigma$ there is an offset vector $o_s \in \mathbb{Z}^X$ and a reset vector $r_s \in \{0,1\}^X$ such that $\rho \xrightarrow{s}_{\mathcal{V}} \rho'$ if and only if $\bigwedge_{x \in X} \rho'(x) = r_s(x)\rho(x) + o_s(x)$. Let $RV(\mathcal{V}, s)$ and $OV(\mathcal{V}, s)$ denote r_s and o_s respectively. Note that $\rightarrow_{\mathcal{V}}$ is uniquely determined by $RV(\mathcal{V}, s)$ and $OV(\mathcal{V}, s)$ for all s.

A labeled semi-linear integer vector addition system with resets (**SVASR**) over variables X and alphabet Σ is a labeled transition system $\mathcal{SV} = \langle \mathbb{Z}^X, \rightarrow_{\mathcal{SV}} \rangle$. For each symbol $s \in \Sigma$ there is an offset semi-linear set $S_s \subseteq \mathbb{Z}^X$ and a reset vector $r_s \in \{0,1\}^X$ such that $\rho \xrightarrow{s}_{\mathcal{SV}} \rho'$ if and only if $\bigwedge_{x \in X} \rho'(x) = r_s(x)\rho(x) + v(x)$ for some $v \in S_s$. Let $RV(\mathcal{SV}, s)$ and $OS(\mathcal{SV}, s)$ denote r_s and S_s respectively. Note that $\rightarrow_{\mathcal{SV}}$ is uniquely determined by $RV(\mathcal{SV}, s)$ and $OS(\mathcal{SV}, s)$ for all s.

A semi-linear transition system over variables X is a labeled transition system $T = \langle \mathbb{Z}^X, \rightarrow_T \rangle$ such that $\xrightarrow{s}_T \subseteq \mathbb{Z}^X \times \mathbb{Z}^X$ is a semi-linear set. This class can also be thought of as the set of transition systems for which transitions are definable in LIA. Since counter machines are semi-linear transition systems, the reachability problem of semi-linear transition systems is undecidable.

A transition formula F over variables X is a linear integer arithmetic formula over free variables X and primed copies X'. For two states $\rho, \rho' \in \mathbb{Z}^X$, we write $[\rho, \rho'] \models F$ if F holds when all $x \in X$ are replaced with $\rho(x)$ and all $x' \in X'$ are replaced with $\rho'(x)$. $TF(X)$ denotes the set of all transition formulas over X.

3 SVASR Reachability Relations in Polynomial Time

The reachability problem for VASRs has been widely studied in the literature. Specifically, it has been shown that the reachability relation $\xrightarrow{L}_{\mathcal{V}}$ is LIA-definable when $L \subseteq \Sigma^*$ is a regular language [3], a communication-free Petri net language

[2], and a context-free language [8]. Given a VASR $\mathcal{V}$ over variables X and alphabet Σ and a language L in the above classes, these works compute a transition formula $F \in TF(X)$ such that $[\rho, \rho'] \models F$ if and only if $\rho \xrightarrow{L}_{\mathcal{V}} \rho'$ in polynomial time. Thus, these algorithms amount to polynomial-time reductions from VASR reachability to satisfiability of existential LIA formulas.

We show that regular reachability of SVASR can be reduced to regular reachability of VASR. Our reduction creates VASR transitions representing the generator representations of the semi-linear sets of the SVASR transitions and encodes the structure of these sets as a regular language over these transitions. A SVASR transition resets some part of the state then adds a vector from a semi-linear set; our key insight is that this is equivalent to a VASR transition applying the same reset and adding one of the base points of the semi-linear set followed by an arbitrary number of VASR transitions adding one of the associated periods.

Consider a SVASR $\mathcal{SV} = \left\langle \mathbb{Z}^X, \rightarrow_{\mathcal{SV}} \right\rangle$ over alphabet Σ and a language $L \subseteq \Sigma^*$. Define a new alphabet $\Sigma_{\mathcal{SV}} \subseteq \{0,1\}^X \times \mathbb{Z}^X$ to be the least set such that:

- For all $s \in \Sigma$, for all $\langle b; P \rangle \in B(OS(\mathcal{SV}, s))$, we have $\langle RV(\mathcal{SV}, s), b \rangle \in \Sigma_{\mathcal{SV}}$
- For all $s \in \Sigma$, for all $\langle b; p_1 \dots p_n \rangle \in B(OS(\mathcal{SV}, s))$, we have $\langle \lambda x.1, p_i \rangle \in \Sigma_{\mathcal{SV}}$ for all $i \in [1, n]$

Define a VASR $\mathcal{V}(\mathcal{SV}) \triangleq \left\langle \mathbb{Z}^X, \rightarrow_{\mathcal{V}(\mathcal{SV})} \right\rangle$ over variables X alphabet $\Sigma_{\mathcal{SV}}$ where

$$RV(\mathcal{V}(\mathcal{SV}), \langle r, v \rangle) = r \qquad OV(\mathcal{V}(\mathcal{SV}), \langle r, v \rangle) = v$$

Finally, for each $s \in S$, define the following regular language $R_s \subseteq \Sigma_{\mathcal{SV}}^*$:

$$R_s \triangleq \bigcup_{\langle b; p_1, \dots, p_n \rangle \in B(OS(\mathcal{SV}, s))} \langle RV(\mathcal{SV}, s), b \rangle \langle \lambda x.1, p_1 \rangle^* \dots \langle \lambda x.1, p_n \rangle^*$$

Lemma 1. *Consider an SVASR $\mathcal{SV}$ over alphabet Σ. For all $s \in \Sigma$, we have:*

$$\left(\rho \xrightarrow{s}_{\mathcal{SV}} \rho'\right) \iff \left(\rho \xrightarrow{R_s}_{\mathcal{V}(\mathcal{SV})} \rho'\right)$$

Proof. ($\implies$) Let X denote the variables of $\mathcal{SV}$. Consider any states ρ, ρ' such that $\rho \xrightarrow{s}_{\mathcal{SV}} \rho'$. We have that $\bigwedge_{x \in X} \rho'(x) = RV(\mathcal{SV}, s)(x)\rho(x) + v(x)$ for some $v \in OS(\mathcal{SV}, s)$. Then, there must be some $\langle b; p_1, \dots, p_n \rangle \in B(OS(\mathcal{SV}, s))$ such that $v = b + \sum_{i=1}^n \lambda_i p_i$ for some $\lambda_1, \dots, \lambda_n \in \mathbb{N}$. Then, observe that the word

$$w \triangleq \langle RV(\mathcal{SV}, s), b \rangle \langle \lambda x.1, p_1 \rangle^{\lambda_1} \dots \langle \lambda x.1, p_n \rangle^{\lambda_n}$$

belongs to R_s and that $\rho \xrightarrow{w}_{\mathcal{V}(\mathcal{SV})} \rho'$, and so $\rho \xrightarrow{R_s}_{\mathcal{V}(\mathcal{SV})} \rho'$.

($\impliedby$) Consider any states ρ, ρ' such that $\rho \xrightarrow{R_s}_{\mathcal{V}(\mathcal{SV})} \rho'$. Then there must be some $w \in R_s$ such that $\rho \xrightarrow{w}_{\mathcal{V}(\mathcal{SV})} \rho'$. By the definition of R_s, for some $\langle b; p_1, \dots, p_n \rangle \in B(OS(\mathcal{SV}, s))$ we have that

$$w = \langle RV(\mathcal{SV}, s), b \rangle \langle \lambda x.1, p_1 \rangle^{\lambda_1} \dots \langle \lambda x.1, p_n \rangle^{\lambda_n}$$

for some $\lambda_1, \ldots, \lambda_n \in \mathbb{N}$. By the definition of $\mathcal{V}(\mathcal{SV})$, this implies that $\bigwedge_{x \in X} \rho'(x) = RV(\mathcal{SV}, s)(x)\rho(x) + (b + \sum_{p \in P} \lambda_p p)$. Then, since $b + \sum_{i=1}^{n} \lambda_i p_i \in OS(\mathcal{SV}, s)$, we have that $\rho \xrightarrow{s}_{\mathcal{SV}} \rho'$.

Then, let $R(L)$ be the language replacing all characters in Σ with their corresponding regular languages: $R(L) \triangleq \{w_0 \ldots w_n : \exists s_0 \ldots s_n \in L, w_i \in R_{s_i}\}$.

Theorem 1. *For language $L \subseteq \Sigma^*$ and semi-linear VASR $\mathcal{SV}$ we have*

$$\left(\rho \xrightarrow{L}_{\mathcal{SV}} \rho'\right) \text{ if and only if } \left(\rho \xrightarrow{R(L)}_{\mathcal{V}(\mathcal{SV})} \rho'\right)$$

Proof. Follows from Lemma 1.

Therefore, L-reachability of SVASR reduces to $R(L)$-reachability of VASR. The extended language $R(L)$ is linearly larger than L with respect to the size of the bases of the semi-linear sets of the SVASR. If L is regular (resp. context-free) (resp. communication-free petri-net language), then $R(L)$ is regular (resp. context-free) (resp. communication-free petri-net language).

4 Best SVASR Abstractions of Semi-linear Transition Systems

We now shift focus to computing over-approximate L-reachability for semi-linear transition systems. This problem has a wide variety of applications, including proving safety properties of computer programs. The following definitions formalize our approach to computing the best SVASR abstraction of a semi-linear transition system for computing over-approximate reachability.

A SVASR-abstraction of labeled transition system T is a pair $\langle f, \mathcal{SV} \rangle$ composed of SVASR $\mathcal{SV}$ and linear simulation f from T to $\mathcal{SV}$. We say that a SVASR-abstraction $\langle f, \mathcal{SV} \rangle$ is a *SVASR-reflection* of T if for any other SVASR-abstraction $\langle f', \mathcal{SV}' \rangle$ of T there is a linear simulation f^* from $\mathcal{SV}$ to $\mathcal{SV}'$ such that $f^* \circ f = f'$.

L-reachability of a SVASR abstraction of T can be used to over-approximate L-reachability of T. A SVASR-reflection $\langle f, \mathcal{SV} \rangle$ is best because its implied over-approximate reachability is at least as precise as any other SVASR abstraction $\langle f', \mathcal{SV}' \rangle$, as $\left\{\langle \rho, \rho' \rangle : f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')\right\} \subseteq \left\{\langle \rho, \rho' \rangle : f^*(f(\rho)) \xrightarrow{L}_{\mathcal{SV}'} f^*(f(\rho'))\right\}$.

This section shows how to compute the SVASR reflection $\langle f, \mathcal{SV} \rangle$ of a semi-linear transition system. We then show how it allows use to define over-approximate L-reachability of semi-linear transition systems in LIA. However, this formula is exponentially sized with respect to the semi-linear transition system, making it impractical for direct usage; we show how to compute a smaller equivalent formula in Sect. 5.

4.1 SVASR-Reflections of Semi-linear Transition Systems

The SVASR reflection $\langle f, \mathcal{SV} \rangle$ of a semi-linear transition system T over variables X and alphabet Σ can be computed as follows. The state space of $\mathcal{SV}$ is $\mathbb{Z}^{X \times \{0,1\}^\Sigma}$. Intuitively, for each variable x and character $s \in \Sigma$, we must make a choice of whether to treat s as a reset of x. Since the choice is arbitrary, we introduce $2^{|\Sigma|}$ copies of each variable x to encode all possible choices. Thus, a "variable" of the SVASR reflection is a pair $\langle x, C \rangle \in X \times \{0,1\}^\Sigma$ where x is variable of T, and $C(s)$ indicates whether s is to be treated as a increment ($C(s) = 1$) or reset ($C(s) = 0$). The simulation $f : \mathbb{Z}^X \to \mathbb{Z}^{X \times \{0,1\}^\Sigma}$ is defined as:

$$f(\rho) = \lambda \langle x, C \rangle . \rho(x)$$

The relation $\to_{\mathcal{SV}}$ is defined in terms of $RV(\mathcal{SV}, s)$ and $OS(\mathcal{SV}, s)$ as:

$$RV(\mathcal{SV}, s) = \lambda \langle x, C \rangle . C(s)$$
$$OS(\mathcal{SV}, s) = \left\{ \lambda \langle x, C \rangle . \text{ if } C(s) \text{ then } \rho'(x) - \rho(x) \text{ else } \rho'(x) : \rho \xrightarrow{s}_T \rho' \right\}$$

A basis for $OS(\mathcal{SV}, s)$ can be computed from a basis for $\xrightarrow{s}_T$ as follows. For any $t = \langle \rho, \rho' \rangle \in \mathbb{Z}^X \times \mathbb{Z}^X$, define

$$\hat{t} \triangleq \lambda \langle x, C \rangle . \text{ if } C(s) \text{ then } \rho'(x) - \rho(x) \text{ else } \rho'(x) \ .$$

Then we define

$$B(OS(\mathcal{SV}, s)) \triangleq \left\{ \left\langle \hat{b}; \hat{p}_1, \ldots, \hat{p}_n \right\rangle : \langle b; p_1, \ldots, p_n \rangle \in B(\xrightarrow{s}_T) \right\} .$$

Lemma 2. *The pair $\langle f, \mathcal{SV} \rangle$ is an abstraction of T.*

Proof. Consider any states ρ, ρ' such that $\rho \xrightarrow{s}_T \rho'$. By the definition of $OS(\mathcal{SV}, s)$, we have that $v = (\lambda \langle x, C \rangle . \text{ if } C(s) \text{ then } \rho'(x) - \rho(x) \text{ else } \rho'(x)) \in OS(\mathcal{SV}, s)$. We can conclude that $f(\rho) \xrightarrow{s}_{\mathcal{SV}} f(\rho')$, since:

$$\bigwedge_{\langle x, C \rangle \in X \times \{0,1\}^\Sigma} f(\rho')(\langle x, C \rangle) = RV(\mathcal{SV}, s)(x, C) f(\rho)(x, C) + v(\langle x, C \rangle)$$
$$\iff \bigwedge_{\langle x, C \rangle \in X \times \{0,1\}^\Sigma, C(s)=1} \rho'(x) = 1\rho(x) + (\rho'(x) - \rho(x))$$
$$\wedge \bigwedge_{\langle x, C \rangle \in X \times \{0,1\}^\Sigma, C(s)=0} \rho'(x) = 0\rho(x) + (\rho'(x))$$

Theorem 2. *The SVASR-abstraction $\langle f, \mathcal{SV} \rangle$ is a SVASR reflection of T.*

Proof. Consider another SVASR abstraction $\langle f', \mathcal{SV}' \rangle$ over variables Y. We will show that the following function $f^* : \mathbb{Z}^{X \times \{0,1\}^\Sigma} \to \mathbb{Z}^Y$ is a simulation from $\mathcal{SV}$ to $\mathcal{SV}'$.

$$f^*(\sigma) = \lambda y . f'(\lambda x . \sigma(\langle x, C_y \rangle))(y) \text{ with } C_y = \lambda s . RV(\mathcal{SV}', s)(y)$$

A piece of intuition for this definition is that $\langle x, C_y\rangle$ is the variable of $\mathcal{SV}$ that abstracts the variable x of T and that experiences the same resets per-character as y in $\mathcal{SV}'$.

First, observe that $f^* \circ f = f'$:

$$\begin{aligned} f^*(f(\rho)) &= \lambda y.f'(\lambda x.f(\rho)(x, C_y))(y) \\ &= \lambda y.f'(\lambda x.\rho(x))(y) = f'(\rho) \end{aligned}$$

To show f^* is a simulation from $\mathcal{SV}$ to $\mathcal{SV}'$, consider states $\sigma, \sigma' \in \mathbb{Z}^{X\times\{0,1\}^\Sigma}$ such that $\sigma \xrightarrow{s}_{\mathcal{SV}} \sigma'$. By the definition of $OS(\mathcal{SV}, s)$, there must be ρ, ρ' such that $\rho \xrightarrow{s}_T \rho'$ and:

$$\bigwedge_{\langle x,C\rangle\in X\times\{0,1\}^\Sigma, C(s)=1} \sigma'(\langle x, C\rangle) = 1\sigma(\langle x, C\rangle) + (\rho'(x) - \rho(x))$$
$$\wedge \bigwedge_{\langle x,C\rangle\in X\times\{0,1\}^\Sigma, C(s)=0} \sigma'(\langle x, C\rangle) = 0\sigma(\langle x, C\rangle) + (\rho'(x))$$

Thus, for all $\langle x, C\rangle \in X \times \{0, 1\}^\Sigma$, if $C(s) = 0$ then $\sigma'(\langle x, C\rangle) = \rho'(x)$ and if $C(s) = 1$ then $\sigma(\langle x, C\rangle) - \sigma'(\langle x, C\rangle) = \rho'(x) - \rho(x)$.

Since $\rho \xrightarrow{s}_T \rho'$, we have that $f'(\rho) \xrightarrow{s}_{\mathcal{SV}'} f'(\rho')$. Then, there is some $v \in OS(\mathcal{SV}', s)$ such that:

$$\bigwedge_{y\in Y} f'(\rho')(y) = RV(\mathcal{SV}', s)(y)f'(\rho)(y) + v(y) \tag{1}$$

Note that $RV(\mathcal{SV}', s)(y) = C_y(s)$ for all s. Then, for all variables $y \in Y$ of SVASR $\mathcal{SV}'$, if $RV(\mathcal{SV}', s)(y) = 0$ then $\sigma'(x, C_y) = \rho'(x)$ by previous reasoning. In such cases:

$$f^*(\sigma')(y) = f'(\lambda x.\sigma'(x, C_y))(y) = f'(\lambda x.\rho'(x))(y) = f'(\rho')(y)$$

Substituting $f'(\rho')(y)$ with $f^*(\sigma')(y)$ in Eq. 1 and using $RV(\mathcal{SV}', s)(y) = 0$, we have that $f^*(\sigma')(y) = RV(\mathcal{SV}', s)(y)f^*(\sigma)(y) + v(y)$.

In the other case, if $RV(\mathcal{SV}, s)(y) = 1$ then $C_y(s) = 1$ and so $\sigma'(x, C_y) - \sigma(x, C_y) = \rho'(x) - \rho(x)$. Then, using the linearity of f', we have:

$$\begin{aligned} f^*(\sigma')(y) - f^*(\sigma)(y) &= f'(\lambda x.\sigma'(x, C_y))(y) - f'(\lambda x.\sigma(x, C_y))(y) \\ &= f'(\lambda x.\sigma'(x, C_y) - \sigma(x, C_y))(y) \\ &= f'(\lambda x.\rho'(x) - \rho(x))(y) = f'(\rho')(y) - f(\rho)(y) \end{aligned}$$

Then, substituting $f'(\rho')(y) - f'(\rho)(y)$ with $f^*(\sigma')(y) - f^*(\sigma)(y)$ in (1), we have that $f^*(\sigma')(y) = RV(\mathcal{SV}', s)(y)f^*(\sigma)(y) + v(y)$.

Therefore, by cases, we have that:

$$\bigwedge_{y\in Y} f^*(\sigma')(y) = RV(\mathcal{SV}', s)(y)f^*(\sigma')(y) + v(y)$$

We can conclude that $f^*(\sigma) \xrightarrow{s}_{\mathcal{SV}'} f^*(\sigma')$, that f^* is a simulation from $\mathcal{SV}$ to $\mathcal{SV}'$, and that $\langle f, \mathcal{SV}\rangle$ is a reflection of T.

4.2 Over-Approximate Semi-linear Transition System Reachability

Using the contents of Sect. 3 and Subsect. 4.1, we have a procedure to compute over-approximate L-reachability of semi-linear transition systems. Given a semi-linear transition system T over variables X, we first compute its SVASR reflection $\langle f, \mathcal{SV}\rangle$ over variables $X \times \{0,1\}^{\Sigma}$. We then compute an over-approximation of the L-reachability relation of T via Sect. 3 and the following lemma:

Lemma 3. *Consider a language $L \subseteq \Sigma^*$, a semi-linear transition system T and its SVASR reflection $\langle f, \mathcal{SV}\rangle$ as defined in Sect. 4. Let $F \in TF(X \times \{0,1\}^{\Sigma})$ be a formula such that $[\sigma, \sigma'] \models F$ if and only if $\sigma \xrightarrow{L}_{\mathcal{SV}} \sigma'$. Define $G \triangleq F[\langle x, C\rangle \mapsto x]$. Then we have*

$$[\rho, \rho'] \models G \text{ if and only if } f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$$

Proof. ($\Longrightarrow$) If $[\rho, \rho'] \models G$ then by the definition of f we have that $[f(\rho), f(\rho')] \models F$ and so $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$. ($\Longleftarrow$) If $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$ then $[f(\rho), f(\rho')] \models F$ and so by the definition of f we have that $[\rho, \rho'] \models G$.

The formula G is an over-approximation of the L-reachability relation of T, as if $\rho \xrightarrow{L}_T \rho'$ then $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$ and so $[\rho, \rho'] \models G$, so it can be used to prove safety properties about T. However, it is too large to be practically useful. The SVASR reflection of a semi-linear transition system has an exponentially larger state space and defining SVASR reachability in LIA takes polynomial space, so the size of G is exponential with respect to the semi-linear transition system.

5 Over-Approximate Semi-linear Transition System Reachability in Polynomial Time

As in Subsect. 4.2, given a semi-linear transition system T, we can compute a formula G such that $[\rho, \rho'] \models G$ if and only if $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$ where $\langle f, \mathcal{SV}\rangle$ is the SVASR-reflection of T. This formula is an over-approximation of the L-reachability relation of T, but requires exponential space w.r.t. T because the dimension of $\mathcal{SV}$ is exponential in the size of the alphabet. We show here that we can compute a formula equivalent to G in polynomial time w.r.t. T.

The key is to never explicitly compute the SVASR-reflection $\langle f, \mathcal{SV}\rangle$. Given a semi-linear transition system T, let $M_T : \Sigma \to (\mathbb{Z}^{X \times \{0,1\}} \times (\mathbb{Z}^{X \times \{0,1\}})^*)^*$ be the function mapping each $s \in \Sigma$ to a basis of the following semilinear set:

$$\left\{\lambda\langle x, r\rangle.\ \text{if } r = 1 \text{ then } \rho'(x) - \rho(x) \text{ else } \rho'(x) : \rho \xrightarrow{s}_T \rho'\right\}$$

A basis of this set can be computed from a basis of $\xrightarrow{s}_T$ in polynomial time, as in Sect. 4.1. Given a semi-linear transition system T over variables X, this subsection computes $G' \in TF(X)$ such that $[\rho, \rho'] \models G'$ if and only if $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$ in polynomial time w.r.t. M_T.

Haase and Halfon [3] recognized that the L-reachability relation of a VASR can be computed via a counting abstraction of L; for each word in L, one must compute the final time that each dimension of the state is reset from left to right and compute the character counts of the subwords in between these final resets. This information is sufficient to compute the composition of VASR transitions along the word because the final reset nullifies the effects of all characters before it, and the effects of all characters after it commute with respect to the counter because they do not reset it; then, their net effect is computable from the character count after the final reset. We consider *abstract trajectories*, a counting abstraction which we will use to identify final resets.

Definition 1. *An* d*-marked* ***abstract trajectory*** *is a function* $n : (\Sigma \times [1, 2d+1]) \to \mathbb{N}$ *such that for all even* i *we have that* $\sum_{s\in\Sigma} n(s,i) \leq 1$.

For a trajectory $w \in \Sigma^*$ and an abstract trajectory n, write $w \Vdash n$ if and only if there exists a decomposition $w = w_1 w_2 \dots w_{2d+1}$ such that for all $i \in [1, 2d+1]$ and $s \in \Sigma$, the value of $n(s,i)$ is the number of times character s appears in subword w_i. The definition ensures that n uniquely identifies w_i for all even i. In this sense, abstract trajectories are a counting abstraction which identifies up to d characters in order from a word and captures the character counts of the subwords in between. We restrict our attention to languages for which the set of abstract trajectories is LIA-definable, and use additional constraints to identify the abstract trajectories that mark the final reset of each dimension.

Formally, we restrict our attention to languages $L \subseteq \Sigma^*$ for which we can compute a formula $AT(L, |\Sigma|)$ over free variables $c_{s,i}$ for all $s \in \Sigma$ and all $i \in \{1 \dots 2|\Sigma| + 1\}$; for any $|\Sigma|$-marked abstract trajectory n, $AT(L, |\Sigma|)$ holds when each $c_{s,i}$ is replaced with $n(s,i)$ if and only if there exists some $w \in L$ such that $w \Vdash n$. Pimpalkhare and Kincaid [8] gave an explicit definition for $AT(L, |\Sigma|)$ in the case that L is context-free; one can adapt techniques from the literature to compute $AT(L, |\Sigma|)$ in the cases that L is regular [3] or a communication-free Petri-net language [2].

Our approach to computing G' is to compute a formula representing the abstract-trajectories of L, constrain the free variables to ensure the final reset of each dimension of the SVASR reflection occurs at an even index, and to encode the resulting SVASR transition for all variables $\langle x, C\rangle$ of the SVASR reflection on the corresponding variable x of T. Directly considering every variable of the SVASR produces an exponentially sized formula. However, the final reset of every SVASR variable will occur at the final occurrence of some character, and at least one SVASR variable for each variable x of the semi-linear transition system will have its final reset at the final occurrence of every character. Our approach is then to mark the final occurrence of each symbol and to conjoin a formula per final occurrence representing the transition of all variables which experience their final reset there.

Leveraging techniques from the literature, we first compute $AT(L, |\Sigma|)$, a formula defining the $|\Sigma|$-marked abstract trajectories of L. We then define a formula $WF(\Sigma)$ ensuring that our symbolic $|\Sigma|$-marked abstract trajectory marks the final occurrence of every $s \in \Sigma$:

$$WF(\Sigma) = \bigwedge_{s \in \Sigma} \left(\bigwedge_{i=1}^{2|\Sigma|+1} \left(\begin{array}{c} c_{s,i} > 0 \implies \\ \bigvee_{\text{even } k \geq i} c_{s,k} > 0 \end{array} \right) \wedge \left(\sum_{j=1}^{|\Sigma|} c_{s,2j} \leq 1 \right) \right)$$

Then, we define the formula $Transition(M_T, \Sigma)$ which describes the transition corresponding to the symbolic abstract trajectory value. This formula uses the following sets of variables to symbolically pick the SVASR translation vector corresponding to each occurrence of $c_{s,i}$:

$$D \triangleq \{d_{s,b,i} : s \in \Sigma, \langle b, P \rangle \in M_T(s), i \in [1, 2|\Sigma| + 1]\}$$

$$E \triangleq \{e_{s,b,p,i} : s \in \Sigma, \langle b, P \rangle \in M_T(s), p \in P, i \in [1, 2|\Sigma| + 1]\}$$

The following formula $Corr(\Sigma, M_T)$ corresponds variables $c_{s,i}$ to the relevant variables $d_{s,b,i}$ and $e_{s,b,p,i}$. The variable $c_{s,i}$ captures how many times s appears in subword i - for each such appearance, we must pick a single base vector and any number of periods.

$$Corr(\Sigma, M_T) \triangleq \bigwedge_{s \in \Sigma} \bigwedge_{i=1}^{2|\Sigma|+1} \left(\sum_{\langle b,P \rangle \in M_T(s)} d_{s,b,i} = c_{s,i} \right) \wedge \bigwedge_{\langle b,P \rangle \in M_T(s)} \bigwedge_{p \in P} (e_{s,b,p,i} > 0 \implies d_{s,b,i} > 0)$$

We define $ResetAt(i, x, M_T, \Sigma)$ to compute the value that x would be reset to by the ith even subword and $AddsAfter(i, x, M_T, \Sigma)$ to compute the value of the increments to x after the ith subword.

$$ResetAt(i, x, M_T, \Sigma) = \sum_{s \in \Sigma} \sum_{\langle b,P \rangle \in M_T(s)} \left(d_{s,b,i} b(\langle x, 0 \rangle) + \sum_{p \in P} e_{s,b,p,i} p(\langle x, 0 \rangle) \right)$$

$$AddsAfter(i, x, M_T, \Sigma) = \sum_{j=i+1}^{2|\Sigma|+1} \sum_{s \in \Sigma} \sum_{\langle b,P \rangle \in M_T(s)} \left(\begin{array}{c} d_{s,b,j} b(\langle x, 1 \rangle) + \\ \sum_{p \in P} e_{s,b,p,j} p(\langle x, 1 \rangle) \end{array} \right)$$

And subsequently define:

$$Transition(X, M_T, \Sigma) \triangleq \bigwedge_{x \in X} \left(\begin{array}{c} x' = x + AddsAfter(0, x, M_T, \Sigma) \wedge \\ \bigwedge_{k=1}^{|\Sigma|} \left(\begin{array}{c} \sum_{s \in \Sigma} c_{s,2k} > 0 \implies \\ x' = ResetAt(2k, x, M_T, \Sigma) + \\ AddsAfter(2k, x, M_T, \Sigma) \end{array} \right) \end{array} \right)$$

Finally, we conjoin these formulae to produce our procedure summary.

$$G'(X, M_T, L, \Sigma) = \begin{array}{c} \exists\{c_{s,i} \geq 0 : s \in \Sigma, i \in [1, 2|\Sigma| + 1]\} \\ \exists\{d_{s,b,i} \geq 0 : d_{s,b,i} \in D\} \exists\{e_{s,b,p,i} \geq 0 : e_{s,b,p,i} \in E\} \\ \left(\begin{array}{c} Transition\,(X, M_T, \Sigma) \wedge AT\,(L, |\Sigma|) \\ \wedge WF\,(\Sigma) \wedge Corr(\Sigma, M_T) \end{array} \right) \end{array}$$

Theorem 3. *Consider a semi-linear transition system T and a language $L \subseteq \Sigma^*$. Let $\langle f, \mathcal{SV}\rangle$ be the SVASR-reflection of T defined in Sect. 4. Then,*

$$[\rho, \rho'] \models G'(X, M_T, L, \Sigma) \iff f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$$

Proof. Firstly, note that there is a one-to-one correspondence between the elements of $S(M_T(s))$ and $OS(\mathcal{SV}, s)$, as both are defined by ρ, ρ' such that $\rho \xrightarrow{s}_T \rho'$. For each $s \in \Sigma$, define $\psi_s : \mathbb{Z}^{X \times \{0,1\}} \to \mathbb{Z}^{X \times \{0,1\}^\Sigma}$ by $\psi_s(v)(x, C) \triangleq v(x, C(s))$ to translate between $S(M_T(s))$ and $OS(\mathcal{SV}, s)$. Let ψ_s^{-1} be a left inverse.

($\implies$) Consider ρ, ρ' such that $[\rho, \rho'] \models G'(X, M_T, L, \Sigma)$. By its definition, there exists a valuation $A : \{c_{s,i} : s \in \Sigma, i \in [1, 2|\Sigma| + 1]\} \cup D \cup E \to \mathbb{N}$ such that $(\mathit{Transition}(X, M_T, \Sigma) \wedge AT(L, |\Sigma|) \wedge \mathit{WF}(\Sigma) \wedge \mathit{Corr}(\Sigma, M_T))$ holds when each $c_{s,i}$ is replaced with $A(c_{s,i})$, each $d_{s,b,i}$ is replaced with $A(d_{s,b,i})$, and each $e_{s,b,p,i}$ is replaced with $A(e_{s,b,p,i})$.

Let $n : (\Sigma \times [1, 2|\Sigma| + 1]) \to \mathbb{N}$ be the function mapping each $\langle s, i\rangle$ to $A(c_{s,i})$. Since this valuation satisfies $AT(L, |\Sigma|)$, we have that n is a $|\Sigma|$-marked abstract trajectory over Σ such that there exists a word $w = s_1 \dots s_{|w|} \in L$ such that $w \Vdash n$. We will show that $f(\rho) \xrightarrow{w}_{\mathcal{SV}} f(\rho')$, therefore showing $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$.

Consider any $\langle x, C\rangle \in X \times \{0,1\}^\Sigma$ such that $RV(\mathcal{SV}, s_i)(\langle x, C\rangle) = 1$ for all s_i in w. Observe that the first conjunct of $Transition(X, M_T, \Sigma)$ ensures that $\rho'(x) = \rho(x) + \sum_{i=1}^{|w|} v_i(\langle x, 1\rangle)$ where each $v_i \in S(M_T(s_i))$. Since it is the case that $v_i(x, 1) = \psi_{s_i}(v_i)(\langle x, C\rangle)$ for all i in $[|w|]$ since $C(s_i) = 1$, we have that $f(\rho')(\langle x, C\rangle) = f(\rho)(\langle x, C\rangle) + \sum_{i=1}^{|w|} \psi_{s_i}(v_i)(\langle x, C\rangle)$.

Consider any $\langle x, C\rangle \in X \times \{0,1\}^\Sigma$ such that $RV(\mathcal{SV}, s_i)(\langle x, C\rangle) = 0$ for some s_i in w. Let j be the highest index such that $RV(\mathcal{SV}, s_j)(\langle x, C\rangle) = 0$. $WF(\Sigma)$ ensures that $A(c_{s_j,2k}) = 1$ for some k. The corresponding conjunct of $\mathit{Transition}(X, M_T, \Sigma)$ ensures that $\rho'(x) = v_j(\langle x, 0\rangle) + \sum_{i=j+1}^{|w|} v_i(\langle x, 1\rangle)$ where each $v_i \in S(M_T(s_i))$. Since it is the case that $v_j(\langle x, 0\rangle) = \psi_{s_j}(v_j)(\langle x, C\rangle)$ since $C(s_j) = 0$ and $v_i(\langle x, 1\rangle) = \psi_{s_i}(v_i)(\langle x, C\rangle)$ for all $i \in [j+1, |w|]$ since $C(s_i) = 1$, we have that $f(\rho')(\langle x, C\rangle) = \psi_{s_j}(v_j)(\langle x, C\rangle) + \sum_{i=j+1}^{|w|} \psi_{s_i}(v_i)(\langle x, C\rangle)$.

Observe that for all $i \in [1, |w|]$, we have $\psi_{s_i}(v_i) \in OS(\mathcal{SV}, s_i)$. Then, by the above casework over all $\langle x, C\rangle$, we have that $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$.

($\impliedby$) Consider ρ, ρ' such that $f(\rho) \xrightarrow{L}_{\mathcal{SV}} f(\rho')$. There exists $w = s_1 \dots s_n \in L$ such that $f(\rho) \xrightarrow{w}_{\mathcal{SV}} f(\rho')$. Let d be the number of unique characters in w and let $i_1 \dots i_d$ be the indexes of the final occurrence of each letter; that is, character s_{i_j} does not appear in subword $s_{i_j+1} \dots s_n$. For all $j \in [1, d]$, let word w_{2j} be the character s_{i_j} and let word w_{2j-1} be the subword $s_{i_{j-1}+1} \dots s_{i_j-1}$; let w_{2d+1} through $w_{2|\Sigma|+1}$ be empty. Let $n : (\Sigma \times [1, 2|\Sigma| + 1]) \to \mathbb{N}$ be the function such that $n(s, i)$ is the number of occurrences of s in w_i. Observe that n is a $|\Sigma|$-marked abstract trajectory and $w \Vdash n$.

By assumption, we have $f(\rho) = \sigma_1 \xrightarrow{s_1}_{\mathcal{SV}} \dots \xrightarrow{s_n} \sigma_{n+1} = f(\rho')$. For all $i \in [1, n]$, let $o_i \in OS(\mathcal{SV}, s_i)$ be the offset vector used in the transition $\sigma_i \xrightarrow{s_i} \sigma_{i+1}$. There exists some $\langle b, P\rangle \in M_T(s_i)$ such that $\psi(o_i) = b + \sum_{p \in P} \lambda_p p$. Fix such a representation for each o_i. Let $\phi : (D \cup E) \to \mathbb{N}$ be the function mapping each $d_{s,b,i}$ to the number of times b occurs in the representations of the o_j corresponding to all s_j in w_i and mapping each $e_{s,b,p,i}$ to the sum of λ_p in the representations of the o_j corresponding to all s_j in w_i.

Finally, we can observe that $G'(X, M_T, L, \Sigma)$ holds when all $c_{s,i}$ are set to $n(s,i)$, all $d \in D$ are set to $\phi(d)$, and all $e \in E$ are set to $\phi(e)$. The subformulas $AT(L, |\Sigma|) \wedge WF(\Sigma)$ and $Corr(\Sigma, M_T)$ hold by construction of n and ϕ respectively. The first conjunct of $Transition(X, M_T, \Sigma)$ holds because the transition $f(\rho) \xrightarrow{w}_{\mathcal{SV}} f(\rho')$ implies that $f(\rho')(x, \lambda s.1) = f(\rho)(x, \lambda s.1) + \sum_{i=1}^{n} o_i(x, \lambda s.1)$ or equivalently $\rho'(x) = \rho(x) + \sum_{i=1}^{n} \psi(o_i)(x, 1)$. For the remainder of the conjuncts, with $k \in [1, d]$ the transition implies that $f(\rho')(x, \lambda s.s = s_{i_k}) = o_{i_k}(x, \lambda s.s = s_{i_k}) + \sum_{j=i_k+1}^{n} o_j(x, \lambda s.s = s_{i_k})$ or equivalently $\rho'(x) = \psi(o_{i_k})(x, 0) + \sum_{j=i_k+1}^{n} \psi(o_j)(x, 1)$. Therefore, $Transition(X, M_T, \Sigma)$ holds, and we therefore have that $G'(X, M_T, L, \Sigma)$ holds.

Observe that $G'(X, M_T, L, \Sigma)$ is polynomially-sized with respect to M_T. We can therefore over-approximate L-reachability of a semi-linear transition system via its exponentially sized SVASR reflection in polynomial time.

6 Related Work

Reachability for Vector Addition Systems over the naturals is decidable [6] but non-elementary [4], prompting the study of integer VAS [3] which operate over integral state vectors. The reachability of integer VAS with resets has been studied widely [2,3,8]; this paper extends such work by considering transition systems in which the set of translation vectors is a potentially infinite semi-linear set. Blondin et al. investigated the extension of integer VAS to affine transformations beyond resets in [1]. Another line [5,9] has investigated extending linear integer arithmetic to include a star operator, effectively computing reachability for Integer Semi-Linear Vector Addition Systems, but not considering resets as we do in this paper. A recent line of work [8,10] has used vector addition systems to compute logical summaries of loops and procedures in computer programs; this work applies a similar recipe with the strictly more powerful domain of SVASRs.

References

1. Blondin, M., Haase, C., Mazowiecki, F., Raskin, M.: Affine extensions of integer vector addition systems with states. Logical Methods Comput. Sci. **17**(3) (2021). https://doi.org/10.46298/lmcs-17(3:1)2021
2. Chistikov, D., Haase, C., Halfon, S.: Context-free commutative grammars with integer counters and resets. Theor. Comput. Sci. **735**, 147–161 (2018). https://doi.org/10.1016/j.tcs.2016.06.017. https://www.sciencedirect.com/science/article/pii/S0304397516302535. Reachability Problems 2014: Special Issue
3. Haase, C., Halfon, S.: Integer vector addition systems with states. In: Ouaknine, J., Potapov, I., Worrell, J. (eds.) RP 2014. LNCS, vol. 8762, pp. 112–124. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-11439-2_9

4. Lazic, R.: The reachability problem for vector addition systems with a stack is not elementary (2013). https://arxiv.org/abs/1310.1767
5. Levatich, M., Bjørner, N., Piskac, R., Shoham, S.: Solving LIA* using approximations. In: Beyer, D., Zufferey, D. (eds.) VMCAI 2020. LNCS, vol. 11990, pp. 360–378. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-39322-9_17
6. Mayr, E.W.: An algorithm for the general Petri net reachability problem. In: Proceedings of the Thirteenth Annual ACM Symposium on Theory of Computing, STOC 1981, pp. 238–246. Association for Computing Machinery, New York (1981). https://doi.org/10.1145/800076.802477
7. Parikh, R.J.: On context-free languages. J. ACM **13**(4), 570–581 (1966). https://doi.org/10.1145/321356.321364
8. Pimpalkhare, N., Kincaid, Z.: Monotone procedure summarization via vector addition systems and inductive potentials. https://nikhilpim.github.io/images/oopsla24.pdf
9. Piskac, R., Kuncak, V.: Linear arithmetic with stars. In: Gupta, A., Malik, S. (eds.) CAV 2008. LNCS, vol. 5123, pp. 268–280. Springer, Heidelberg (2008). https://doi.org/10.1007/978-3-540-70545-1_25
10. Silverman, J., Kincaid, Z.: Loop summarization with rational vector addition systems. In: Dillig, I., Tasiran, S. (eds.) CAV 2019. LNCS, vol. 11562, pp. 97–115. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-25543-5_7