Minimizing Cost Register Automata over a Field

Yahia Idriss Benalioua
Aix Marseille Univ, CNRS, LIS, Marseille, France

Nathan Lhote
Aix Marseille Univ, CNRS, LIS, Marseille, France

Pierre-Alain Reynier
Aix Marseille Univ, CNRS, LIS, Marseille, France

Abstract

Weighted automata (WA) are an extension of finite automata that define functions from words to values in a given semiring. An alternative deterministic model, called Cost Register Automata (CRA), was introduced by Alur et al. It enriches deterministic finite automata with a finite number of registers, which store values, updated at each transition using the operations of the semiring. It is known that CRA with register updates defined by linear maps have the same expressiveness as WA. Previous works have studied the register minimization problem: given a function computable by a WA and an integer k, is it possible to realize it using a CRA with at most k registers?

In this paper, we solve this problem for CRA over a field with linear register updates, using the notion of linear hull, an algebraic invariant of WA introduced recently by Bell and Smertnig. We then generalise the approach to solve a more challenging problem, that consists in minimizing simultaneously the number of states and that of registers. In addition, we also lift our results to the setting of CRA with affine updates. Last, while the linear hull was recently shown to be computable by Bell and Smertnig, no complexity bounds were given. To fill this gap, we provide two new algorithms to compute invariants of WA. This allows us to show that the register (resp. state-register) minimization problem can be solved in 2-ExpTime (resp. in NExpTime).

2012 ACM Subject Classification Theory of computation → Quantitative automata

Keywords and phrases Weighted automata, Cost Register automata, Zariski topology

Digital Object Identifier 10.4230/LIPIcs.MFCS.2024.23

Related Version *Full Version*: https://arxiv.org/abs/2307.13505 [5]

Funding This work has been partly funded by the QuaSy project (ANR-23-CE48-0008).

1 Introduction

Weighted automata (WA). WA are a quantitative extension of finite state automata and have been studied since the sixties [17]. These automata define functions from words to a given semiring: each transition has a weight in the semiring and the weight of an execution is the product of the weights of the transitions therein; the non-determinism of the model is handled using the sum of the semiring: the weight associated with a word is the sum of the weights of the different executions over this word. Functions realized by weighted automata are called rational series. This fundamental model has been widely studied during the last decades [14]. While some expressiveness results can be obtained in a general framework (such as the equivalence with rational expressions), the decidability status of important problems heavily depends on the considered semiring. Amongst the classical problems of interest, one can mention *equivalence*, *sequentiality* (resp. *unambiguity*), which aims at determining whether there exists an equivalent deterministic (resp. unambiguous) WA, and *minimization*, which aims at minimizing the number of states.

Weighted automata over a field (e.g. the field of rationals $\mathbb{Q}$) enjoy many nice properties: the equivalence of weighted automata is decidable and they can be minimized, and both can be done efficiently (see e.g. [16, Theorem 4.10 and Corollary 4.17 (Chapter III)]). The

49th International Symposium on Mathematical Foundations of Computer Science (MFCS 2024).
Editors: Rastislav Královič and Antonín Kučera; Article No. 23; pp. 23:1–23:15
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

sequentiality and unambiguity are also decidable, as shown recently in [3, 4], with no complexity bounds however. The most studied semirings which are not fields are the tropical semirings and the semiring of languages, and in both cases equivalence is undecidable (see [9, Section 3] and [6, Theorem 8.4]) and no minimization algorithm is known. Regarding sequentiality, partial decidability results have been obtained for these semirings using the notion of twinning property [7, 15].

Cost register automata (CRA). CRA have been introduced more recently by Alur et al. [1]. A cost register automaton is a deterministic finite state automaton endowed with a finite number of registers storing values from the semiring. The registers are initialized by some values, then at each transition the values are updated using the operations and constants of the semiring. Several fragments of CRA can be considered by restricting the operations allowed. For instance, an easy observation is that WA are exactly CRA with one state (however, one can observe that adding states does not extend expressiveness) and linear updates, *i.e.* updates of the form $X \coloneqq \sum_{i=1}^{k} X_i * c_i$ (intuitively, the new values of the registers only depend linearly on the previous ones). Thus, the model of linear CRA is an alternative to WA which allows to trade non-determinism for registers.

The register minimization problem. As CRA are finite state automata extended with registers storing elements from the semiring, it is natural to aim at minimizing the number of registers used. For a given class $\mathcal{C}$ of CRA, this problem asks, given a WA and a number k, whether there exists an equivalent CRA in $\mathcal{C}$ with at most k registers. From a practical point of view, reducing the number of registers allows to reduce the memory usage, since a register can require unbounded memory. From a theoretical point of view, this problem can be understood as a refinement of the classical problem of minimization of WA. Indeed, a WA can be translated into a linear CRA with a single state, and as many registers as the number of states of the WA. This problem has been studied in [2, 11, 10] for three different models of CRA but in all these works, the additive law of the semiring is not allowed (*i.e.* updates of the form $X \coloneqq Y + Z$ are forbidden). It is worth noticing that [11] encompasses the case of CRA over a field, with only updates of the form $X \coloneqq Y * c$, with c an element of the field.

While the minimal number of registers needed to realise a WA (also known as the *register complexity*) is upper bounded by the number of states of a minimal WA, it may be possible to build an equivalent CRA with fewer registers, but more states. Hence there is a *tradeoff* between the number of states and the number of registers. This leads to the following *state-register minimization problem for* CRA which asks, for a class $\mathcal{C}$ of CRA, given a WA and integers n, k whether an equivalent CRA in $\mathcal{C}$ with n states and k registers can be constructed. In this framework, the classical minimization of WA corresponds to minimizing the number of registers while using only one state, for the class of linear CRA.

The linear hull. As mentioned before, the case of fields is well-behaved to obtain decidability results. In their recent work [3], Bell and Smertnig introduced the notion of *linear hull* of a WA. This notion is inspired by the algebraic theory needed to study polynomial automata but cast into a linear setting. A linear algebraic set (aka linear Zariski closed set) is a finite union of vector subspaces: we later call them *Z-linear sets*. Given a Z-linear set $S = \bigcup_{i=1}^{p} V_i$, the dimension of S is the maximum of the dimensions of the V_is. In this work, the size of the union, p, is called the length of S. Observe that such Z-linear sets were also used in [8] for a category-theoretic approach to minimization of weighted automata over a field. We say such a set is an invariant if it contains the initial vector and is stable under the updates of the automaton. Then the linear hull of a weighted automaton is the strongest Z-linear invariant. In [3], Bell & Smertnig show that computing the linear hull of a minimal automaton allows to decide sequentiality and unambiguity. In addition, in [4], they show that the linear hull can effectively be computed, without providing complexity bounds however.

Contributions. In this work, we deepen the analysis of the linear hull of a WA in order to solve the register and state-register minimization problems for linear CRA. In addition, we also provide new algorithms to compute the linear hull which come with complexity upper bounds, which can be used to derive complexity results for minimization problems as well as for sequentiality and unambiguity of WA. More precisely, our contributions are as follows:

- Firstly, we show that *the register minimization problem* for the class of linear CRA over a field can be solved in 2-ExpTime. To this end, given a rational series f, we show that the minimal number of registers needed to realize f using a linear CRA is exactly the dimension of the linear hull of a minimal WA of f. We then show that the linear hull of a WA can be computed in 2-ExpTime. We show that this complexity drops down to ExpTime for the particular case of commuting transition matrices (which includes the case of a single letter alphabet), with a matching lower bound.
- As a consequence of the computation of the linear hull of a WA and of results proved in [3], we obtain a 2-ExpTime upper bound for the problems of *sequentiality and unambiguity of weighted automata* over a field, closing a question raised in [4].
- Secondly, we prove that the *state-register minimization problem* for linear CRA can be solved in NExpTime. More precisely, given a minimal WA A, we show a correspondence between Z-linear invariants of A and linear CRA equivalent to A. This correspondence maps the length (resp. dimension) of the invariant to the number of states (resp. registers) of the equivalent linear CRA. We then provide a (constructive) NExpTime algorithm that, given a minimal WA and two integers n, k, guesses a well-behaved invariant allowing to exhibit a satisfying equivalent CRA.
- Last, we actually present these results in a more general setting, by considering *affine* CRA, which are a slight extension of linear CRA allowing to use affine maps in the updates of the registers.

Outline of the paper. We present the models of weighted automata and cost register automata in Section 2. We then formally define the two problems we consider, *i.e.* register and state-register minimization problems, and state our main results in Section 3. In Section 4, we introduce the necessary topological notions to define Z-linear/Z-affine set and invariants of weighted automata, and detail our characterizations of the register and state-register complexities of a rational series. Finally in Section 5, we present our algorithms, as well as their consequences in terms of decidability and complexity for the two problems we consider. Omitted proofs and more details for Sections 4 and 5 can be found in the appendix of the full version of this paper [5].

2 Weighted Automata and Cost Register Automata

Basic concepts and notations. An alphabet Σ is a finite set of letters. The set of finite words over Σ will be denoted by Σ^*, the empty word by ϵ and, for two words u and v, uv will denote their concatenation. For two sets X and Y, we denote by $X \times Y$ their cartesian product and by $\pi_X : X \times Y \to X$ and $\pi_Y : X \times Y \to Y$ we denote the canonical projection on X and Y respectively. The set nonnegative integers will be denoted by $\mathbb{N}$. For two integers i, j, we will denote by $[\![i, j]\!]$ the interval of integers between i and j (both included).

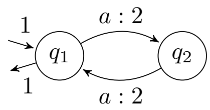

Figure 1 The WA of Example 2.

A *semigroup* $(S, *)$ is a set S together with an associative binary operation $*$. If $(S, *)$ has an identity element e, $(S, *, e)$ is called a *monoid* and if, moreover, every element has an inverse, $(S, *, e)$ is called a *group*. If there is no ambiguity, we will identify algebraic structures with the set that they are defined on. A semigroup (or a monoid/group) is said to be *commutative* if its law is. A sub-semigroup (or submonoid/subgroup) of S is a subset of S that is a semigroup (or a monoid/group). Given $E \subseteq S$, the monoid *generated* by E, denoted $\langle E \rangle$, is the smallest sub-monoid of S containing E.

A *field* $(\mathbb{K}, +, \cdot)$ is a structure where $(\mathbb{K}, +, 0)$ and $(\mathbb{K} \setminus \{0\}, \cdot, 1)$ are commutative groups and multiplication distributes over addition. In this work, we will consider $\mathbb{K}$ as the field of rational numbers $\mathbb{Q}$, or any finite field extension of $\mathbb{Q}$, to perform basic operations in polynomial time. For all $n \in \mathbb{N}$, $\mathbb{K}^n$ is an n-dimensional *vector space* over the field $\mathbb{K}$. We will work with row vectors and apply matrices on the right, and we will identify linear maps (resp. linear forms) with their corresponding matrices (resp. column vectors). The set of n by m matrices over $\mathbb{K}$ will be denoted by $\mathbb{K}^{n \times m}$, and $\mathbb{K}^{1 \times n}$ (or simply $\mathbb{K}^n$ when there is no ambiguity) will denote the set of n-dimensional vectors. For any matrix M (resp. vector v), and indices i and j, $M_{i,j}$ (resp. v_i) will denote the value of the entry in the i-th row and the j-th column of M (resp. the i-th entry of v). Matrix transposition will be denoted by M^t. A *vector subspace* of $\mathbb{K}^n$ is a subset of $\mathbb{K}^n$ stable by linear combinations and for all subsets E of $\mathbb{K}^n$, $\mathsf{span}(E)$ will denote the smallest vector subspace of $\mathbb{K}^n$ containing E (if E contains a single vector $(x_1, \ldots, x_n)$, $\mathsf{span}(E)$ will be denoted by $\mathsf{span}(x_1, \ldots, x_n)$).

$\mathbb{K}^n$ can also be seen as an n-dimensional *affine space*. Affine maps $f : \mathbb{K}^n \to \mathbb{K}^m$ are maps of the form $f(u) = uf^{(l)} + f^{(a)}$ where $f^{(l)} \in \mathbb{K}^{n \times m}$ and $f^{(a)} \in \mathbb{K}^{1 \times m}$. An *affine subspace* A of $\mathbb{K}^n$ is a subset of $\mathbb{K}^n$ of the form $A = p + V$ with $p \in A$ and V a vector subspace of $\mathbb{K}^n$. They are stable by affine combinations (linear combinations with coefficients adding up to 1). For all $E \subseteq \mathbb{K}^n$, $\mathsf{aff}(E)$ will denote the smallest affine subspace of $\mathbb{K}^n$ containing E.

Weighted Automata. Let Σ be a finite alphabet and $(\mathbb{K}, +, \cdot)$ be a field.

▶ **Definition 1** (Weighted Automaton). *A* Weighted Automaton *(*WA *for short) of dimension d, on Σ over $\mathbb{K}$, is a triple $\mathcal{R} = (u, \mu, v)$, where $u \in \mathbb{K}^{1 \times d}$, $v \in \mathbb{K}^{d \times 1}$ and $\mu : \Sigma^* \to \mathbb{K}^{d \times d}$ is a monoid morphism. We will call u and v the* initial *and* terminal *vectors respectively and $\mu(a)$, for $a \in \Sigma$, will be called a* transition matrix. *A* WA *realizes a* formal power series *over Σ^* with coefficients in $\mathbb{K}$ (a function from Σ^* to $\mathbb{K}$) defined, for all $w \in \Sigma^*$, by $[\![\mathcal{R}]\!](w) = u\mu(w)v$. Any series that can be realized by a* WA *will be called* rational.

WA also have a representation in terms of finite-state automata, in which transitions are equipped with weights. We then say that a WA is sequential (resp. unambiguous) when its underlying automaton is. Formally, we say that a WA $\mathcal{R} = (u, \mu, v)$ is *sequential* when u has a single non-zero entry and, for each letter a, and each index i, there is at most one index j such that $\mu(a)_{i,j} \neq 0$.

▶ **Example 2.** We consider the WA, on the alphabet $\{a\}$ and over the field of real numbers, $\mathcal{R} = (u, \mu, v)$ with $u = (1, 0)$, $v = (1, 0)^t$, and $\mu(a) = \begin{pmatrix} 0 & 2 \\ 2 & 0 \end{pmatrix}$. One can verify that the function realized by this WA maps the word a^n to 2^n if n is even, and to 0 otherwise. It can be represented graphically by the automaton depicted on Figure 1.

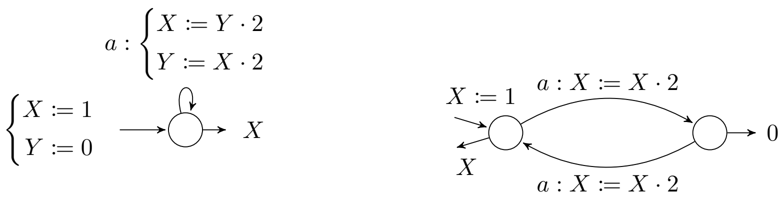

Figure 2 Two CRA detailed in Example 5. Registers are denoted by letters X, Y.

A WA realizing a rational series f is said to be *minimal* if its dimension is minimal among all the WA realizing f. We also have the following characterization of minimal WA (see [16, Proposition 4.8 (Chapter III)]):

▶ **Proposition 3.** *Let $\mathcal{R} = (u, \mu, v)$ be a d-dimensional* WA *and let* $\mathsf{LR}(\mathcal{R}) = u\mu(\Sigma^*) = \{u\mu(w) \mid w \in \Sigma^*\}$ *be its (left)* reachability set *and* $\mathsf{RR}(\mathcal{R}) = \mu(\Sigma^*)v$ *be its right reachability set.*

$\mathcal{R}$ *is a minimal* WA *if and only if* $\mathsf{span}(\mathsf{LR}(\mathcal{R})) = \mathbb{K}^{1\times d}$ *and* $\mathsf{span}(\mathsf{RR}(\mathcal{R})) = \mathbb{K}^{d\times 1}$*.*

Expressions, substitutions and Cost Register Automata. For a field $(\mathbb{K}, +, \cdot)$ and a finite set of variables $\mathcal{X}$ disjoint from $\mathbb{K}$, let $\mathsf{Exp}(\mathcal{X})$ denote the set of expressions generated by the grammar $e ::= k \mid X \mid e + e \mid e \cdot e$, where $k \in \mathbb{K}$ and $X \in \mathcal{X}$. A *substitution* over $\mathcal{X}$ is a map $s \colon \mathcal{X} \to \mathsf{Exp}(\mathcal{X})$. It can be extended to a map $\mathsf{Exp}(\mathcal{X}) \to \mathsf{Exp}(\mathcal{X})$ by substituting each variable X in the expression given as an input by $s(X)$. By identifying s with its extension, we can compose substitutions. We call *valuations* the substitutions of the form $v \colon \mathcal{X} \to \mathbb{K}$. The set of substitutions over $\mathcal{X}$ will be denoted by $\mathsf{Sub}(\mathcal{X})$ and the set of valuations $\mathsf{Val}(\mathcal{X})$.

▶ **Definition 4** (Cost Register Automaton). *A* cost register automaton *(*CRA *for short), on the alphabet Σ over the field $\mathbb{K}$, is a tuple $\mathcal{A} = (Q, q_0, \mathcal{X}, v_0, o, \delta)$ where Q is a finite set of* states*, $q_0 \in Q$ is the initial state, $\mathcal{X}$ is a finite set of registers (variables), $v_0 \in \mathsf{Val}(\mathcal{X})$ is the registers' initial valuation, $o \colon Q \to \mathsf{Exp}(\mathcal{X})$ is the output function, and $\delta \colon Q \times \Sigma \to Q \times \mathsf{Sub}(\mathcal{X})$ is the transition function. We will denote by $\delta_Q := \pi_Q \circ \delta$ the transition function of the underlying automaton of the* CRA *and $\delta_{\mathcal{X}} := \pi_{\mathsf{Sub}(\mathcal{X})} \circ \delta$ its register update function.*

$\mathcal{A}$ computes a function $[\![\mathcal{A}]\!] \colon \Sigma^* \to \mathbb{K}$ defined as follows: the configurations of $\mathcal{A}$ are pairs $(q, v) \in Q \times \mathsf{Val}(\mathcal{X})$. The run of $\mathcal{A}$ on a word $w = a_1 \dots a_n \in \Sigma^*$ is the sequence of configurations $(q_i, v_i)_{i \in [\![0,n]\!]}$ where, q_0 is the initial state, v_0 is the initial valuation and, for all $i \in [\![1, n]\!]$, $q_i = \delta_Q(q_{i-1}, a_i)$ and $v_i = v_{i-1} \circ \delta_{\mathcal{X}}(q_{i-1}, a_i)$. We then define $[\![\mathcal{A}]\!](w) = v_n(o(q_n))$.

δ can be extended to words by setting, for all $q \in Q$, $\delta(q, \epsilon) = (q, id_{\mathcal{X}})$, where $id_{\mathcal{X}}$ is the substitution such that $id_{\mathcal{X}}(X) = X$ for all $X \in \mathcal{X}$, and, for all $a \in \Sigma$ and $w \in \Sigma^*$, $\delta_Q(q, aw) = \delta_Q(\delta_Q(q, a), w)$ and $\delta_{\mathcal{X}}(q, aw) = \delta_{\mathcal{X}}(q, a) \circ \delta_{\mathcal{X}}(\delta_Q(q, a), w)$. We then have

$$[\![\mathcal{A}]\!](w) = v_0 \circ \delta_{\mathcal{X}}(q_0, w)(o(\delta_Q(q_0, w)))$$

▶ **Example 5** (Example 2 continued). Two CRA are depicted on Figure 2. They are both on the alphabet $\{a\}$ and over the field of real numbers, and both realize the same function as the WA considered in Example 2.

An expression is called *linear* if it has the form $\sum_{i=1}^{k} \alpha_i X_i$, for some family of $\alpha_i \in \mathbb{K}$ and $X_i \in \mathcal{X}$, and if it has the form $\sum_{i=1}^{k} \alpha_i X_i + \beta$, for some $\beta \in \mathbb{K}$, it is called *affine*. We will denote by $\mathsf{Exp}_\ell(\mathcal{X})$ (resp. $\mathsf{Exp}_a(\mathcal{X})$) the set of linear (resp. affine) expressions.

▶ **Definition 6** (Linear/Affine CRA). *A* CRA $\mathcal{A} = (Q, q_0, \mathcal{X}, v_0, o, \delta)$ *is called* linear *if,* $\delta_{\mathcal{X}}(q, a)(X) \in \mathsf{Exp}_\ell(\mathcal{X})$ *and* $o(q) \in \mathsf{Exp}_\ell(\mathcal{X})$*, for all* $q \in Q, a \in \Sigma$ *and* $X \in \mathcal{X}$*, and if* $\delta_{\mathcal{X}}(q, a)(X) \in \mathsf{Exp}_a(\mathcal{X})$ *and* $o(q) \in \mathsf{Exp}_a(\mathcal{X})$*, the* CRA *is called* affine*.*

Linear CRA are a particular case of affine CRA and, given an affine CRA it is always possible to define an equivalent linear CRA using one more register with a constant value of 1 to realize affine register updates in a linear way, thus :

▶ Remark 7. Linear and affine CRA have the same expressiveness.

The added cost of a register will however become relevant when we will consider minimization problems in the next sections.

Observe that we can assume that $\mathcal{X} = \{X_1, \ldots, X_k\}$ is ordered, and identify any linear expression $e = \sum_{i=1}^{k} \alpha_i X_i$ (with the α_i not present in the expression assumed to be 0) with the linear form $\underline{e} \colon \mathbb{K}^k \to \mathbb{K}$ defined by the column vector $(\alpha_1, \ldots, \alpha_k)^t$. We can then identify any linear substitution $s \colon \mathcal{X} \to \mathsf{Exp}_\ell(\mathcal{X})$ with the linear map $\underline{s} \colon \mathbb{K}^k \to \mathbb{K}^k$ defined by the block matrix $(\underline{s(X_1)}|\cdots|\underline{s(X_k)})$, and we can identify any valuation $v \colon \mathcal{X} \to \mathbb{K}$ with the vector $\underline{v} = (v(X_1), \cdots, v(X_k))$ of the vector space $\mathbb{K}^k$.

In the following, we will drop the underline notation and make the identifications implicitly.

Thanks to these observations, the registers of a linear CRA and their updates can be characterized by the values of the vector associated with v_0, and the linear maps associated with the $\delta_{\mathcal{X}}(q, a)$ and $o(q)$, for all $q \in Q$ and $a \in \Sigma$, and we can check that

$$[\![\mathcal{A}]\!](w) = v_0 \; \delta_{\mathcal{X}}(q_0, w) \; o(\delta_Q(q_0, w))$$

We can also identify affine expressions with affine forms and affine substitutions with affine maps to simplify dealing with affine CRA. We define and use these identifications in Appendix A.2 of the full version of this paper [5].

▶ **Proposition 8** ([1]). *There is a bijection between* WA *and linear* CRA *with a single state.*

Given a WA, one can build an equivalent CRA with as many registers as states of the WA: for each letter a, the transition matrix $\mu(a)$ can be interpreted as a (linear) substitution, associated with the self-loop of label a. The converse easily follows from the previous observations when the CRA has a single state.

▶ **Example 9** (Example 2 continued). The CRA depicted on the left of Figure 2 is obtained by the translation of the WA of Figure 1 into CRA with a single state.

▶ Remark 10. Sequential WA are exactly linear CRA with a single register.

Indeed, both sequential WA and linear CRA with only one register are deterministic finite automata that can also store a single value updated at each transition using only products. They can then be identified.

3 Problems and Main Results

▶ **Definition 11** (Register minimization problem). *Given a class* $\mathcal{C}$ *of* CRA*, we ask:*

- ***Input:*** *a rational series* f *realized by a given* WA*, and an integer* $k \in \mathbb{N}$
- ***Question:*** *Does there exist a* CRA *realizing* f *in the class* $\mathcal{C}$ *with at most* k *registers?*

We will show this problem is decidable for the classes of linear and affine CRA:

▶ **Theorem 12.** *The register minimization problem is decidable for the classes of linear and affine* CRA *in 2-*EXPTIME*. Furthermore, the algorithm exhibits a solution when it exists.*

For a rational series f, the minimal number of registers needed to realize f using CRA in some class $\mathcal{C}$ is called its *register complexity with respect to class* $\mathcal{C}$. Dually, if one wants to minimize the number of states, then we know we can always build a linear (hence affine) CRA with a single state (Proposition 8). A more ambitious goal is to try to reduce simultaneously the number of states and of registers, in some given class $\mathcal{C}$ of CRA. Observe that, in general, there is no CRA with minimal numbers of both states and registers (see Example 5). Given a rational series f, we say that a pair (n, k) is *optimal* if f can be realized by a CRA in class $\mathcal{C}$ with n states and k registers and no CRA of $\mathcal{C}$ realizing f with at most n states can have strictly less than k registers and vice-versa.

Formally, we call the *state-register complexity with respect to class* $\mathcal{C}$ of a rational series f, the set of optimal pairs of integers (n, k).

This leads to the definition of a second minimization problem:

▶ **Definition 13** (State-Register minimization problem)**.** *Given a class* $\mathcal{C}$ *of* CRA*, we ask:*

- ***Input:*** *a rational series* f *realized by a given* WA*, and two integers* $n, k \in \mathbb{N}$
- ***Question:*** *Does there exist a* CRA *realizing* f *in the class* $\mathcal{C}$ *with at most* n *states and at most* k *registers?*

In the sequel, we solve this problem for linear and affine CRA:

▶ **Theorem 14.** *The state-register minimization problem is decidable for the classes of linear and affine* CRA *in NExpTime. Furthermore, the algorithm exhibits a solution when it exists.*

▶ Remark 15. The complexities we give are valid for fields where it is possible to perform elementary operations efficiently (e.g. $\mathbb{Q}$). See Remark 40 for a more detailed discussion on the matter.

4 Characterizing the state-register complexity using invariants of WA

4.1 Zariski topologies and invariants of WA

Let $\mathbb{K}$ be a field and $n \in \mathbb{N}$. The *Zariski topology* on $\mathbb{K}^n$ is defined as the topology whose closed sets are the sets of common roots of a finite collection of polynomials of $\mathbb{K}[X_1, \ldots, X_n]$. A linear version of this topology, called the *linear Zariski topology*, was introduced by Bell and Smertnig in [3]. Its closed sets, which we will call *Z-linear sets*, are finite unions of vector subspaces of $\mathbb{K}^n$.

A set $S \subseteq \mathbb{K}^n$ is called *irreducible* if, for all closed sets C_1 and C_2, such that $S \subseteq C_1 \cup C_2$, we have either $S \subseteq C_1$ or $S \subseteq C_2$. The Zariski topologies defined above are Noetherian topologies in which every closed set can be written as a finite union of irreducible components. We then define the *dimension* of a Z-linear set as the maximum dimension of its irreducible components and their number will be called its *length*.

For a set $S \subseteq \mathbb{K}^n$, $\overline{S}^\ell$ will denote its closure in the linear Zariski topology. In this topology, closed irreducible sets are vector subspaces of $\mathbb{K}^n$ and linear maps are continuous and closed maps (mapping closed sets to closed sets). In particular, for all $S \subseteq \mathbb{K}^n$ and linear map $f : \mathbb{K}^n \to \mathbb{K}^n$,$\overline{f(S)}^\ell = f(\overline{S}^\ell)$. Moreover, if $S \subseteq \mathbb{K}^n$ is irreducible and $f : \mathbb{K}^n \to \mathbb{K}^n$ is continuous, then $f(S)$ is irreducible. These properties will be used implicitly in the following (see [3, Lemma 3.5] for more details and references).

We will also define an affine version of this topology that enjoy the same properties in Subsection 4.4.

▶ **Definition 16.** *Let Σ be a finite alphabet and let $\mathcal{R} = (u, \mu, v)$ be a d-dimensional* WA *on Σ over $\mathbb{K}$. A subset $I \subseteq \mathbb{K}^d$ is called an* invariant *of $\mathcal{R}$ if $u \in I$ and, for all $w \in I$ and $a \in \Sigma$, $w\mu(a) \in I$. For two invariants I_1 and I_2, we say that I_1 is* stronger *than I_2 if $I_1 \subseteq I_2$. In particular, the strongest invariant of $\mathcal{R}$ is its* reachability set $\mathsf{LR}(\mathcal{R}) = u\mu(\Sigma^*)$.

An invariant that is also a Z-linear set will be called a Z-linear *invariant. The strongest Z-linear invariant of $\mathcal{R}$ is the closure of $\mathsf{LR}(\mathcal{R})$ in the linear Zariski topology (which is well-defined since the topology is Noetherian).*

▶ **Example 17** (Example 2 continued). The reachability set of the WA considered in Example 2 is $\mathsf{LR}(\mathcal{R}) = \{(2^{2n}, 0) \mid n \in \mathbb{N}\} \cup \{(0, 2^{2n+1}) \mid n \in \mathbb{N}\}$. Its strongest Z-linear invariant is then the union of the two coordinate axes of the plane $\overline{\mathsf{LR}(\mathcal{R})}^{\ell} = \mathsf{span}(1,0) \cup \mathsf{span}(0,1)$.

Indeed, the inclusion $\subseteq$ comes from the fact that $u = (1,0) \in \mathsf{span}(1,0)$ and $\mathsf{span}(1,0) \cup \mathsf{span}(0,1)$ is stable by multiplication by $\mu(a)$ and the inclusion $\supseteq$ comes from the fact that, for the linear Zariski topology, $\{(1,0)\}$ is dense in $\mathsf{span}(1,0)$ and $\{(0,2)\}$ is dense in $\mathsf{span}(0,1)$.

▶ Remark 18. In the previous example, the strongest Z-linear invariant is actually the strongest algebraic invariant (*i.e.* closed in the Zariski topology). Of course, this is not always the case.

The Z-linear invariants of two WA realizing the same function do not necessarily coincide but, since $\mathbb{K}$ is a field, it is well-known that for every rational series f, there exists a (computable) minimal WA realizing f that is unique up to similarity in the following sense (see [16, Proposition 4.10 (Chapter III)]):

▶ **Definition 19.** *Let $\mathcal{R} = (u, \mu, v)$ and $\mathcal{R}' = (u', \mu', v')$ be two d-dimensional* WA *over $\mathbb{K}$.*

$\mathcal{R}$ and $\mathcal{R}'$ are said to be similar *if there exists an invertible (change of basis) matrix $P \in \mathbb{K}^{d \times d}$ such that $u' = uP$, $\mu'(a) = P^{-1}\mu(a)P$ for all $a \in \Sigma$ and $v' = P^{-1}v$.*

▶ Remark 20. The Z-linear invariants of two similar WA $\mathcal{R}$ and $\mathcal{R}'$ only differ by a change of basis. i.e. there is a bijection between the Z-linear invariants of $\mathcal{R}$ and those of $\mathcal{R}'$ that, in particular, preserves the length and dimension.

4.2 Strongest invariants and characterization

The notion of strongest Z-linear invariant was introduced by Bell and Smertnig in [3], under the name “linear hull”. They showed, in [4], that it is computable and can be used to decide whether a WA is equivalent to a deterministic (or an unambiguous) one.

▶ **Theorem 21** ([3, Theorem 1.3]). *A rational series f can be realized by a sequential* WA *iff the strongest Z-linear invariant of a minimal* WA *realizing f has dimension at most 1.*

The following result generalizes this theorem by linking linear CRA to Z-linear invariants. It constitutes the key characterization that will allow us to solve the minimization problems.

▶ **Theorem 22** (Characterization). *Let f be a rational series. Then f can be realized by a linear* CRA *with n states and k registers iff there exists a minimal* WA *realizing f that has a Z-linear invariant of length at most n and dimension at most k.*

As we will see in Subsection 4.4, this theorem can also be extended to affine CRA.

Observe that, thanks to Remark 20, the property of the above characterization is actually valid for *every* minimal WA realizing f. Moreover, since the dimension of the strongest Z-linear invariant is minimal, finding this dimension allows to solve the register minimization problem for linear CRA. This is formalized in the following result, which generalizes Theorem 21 thanks to Remark 10.

▶ **Corollary 23.** *The register complexity of a rational series f w.r.t. the class of linear* CRA *is the dimension of the strongest Z-linear invariant of any minimal* WA *realizing f.*

An immediate consequence of this result is that computing the strongest invariant allows to decide the register minimization problem.

▶ **Example 24** (Example 2 continued)**.** As we have seen in Example 17, $\overline{\mathsf{LR}(\mathcal{R})}^{\ell}$ is 1-dimensional and has two irreducible components, thus $[\![\mathcal{R}]\!]$ can be realized by a CRA with two states and one register (depicted on the right of Figure 2).

4.3 Invariants of minimal WA and correspondence with CRA

▶ **Proposition 25.** *Let $\mathcal{R}$ be a* WA *realizing a rational series f. If $\mathcal{R}$ has a Z-linear invariant of length n and dimension k, then every minimal* WA *realizing f has a Z-linear invariant of length $\leq n$ and dimension $\leq k$.*

Thanks to Remark 20, it suffices to show the existence of one minimal WA verifying the proposition, since they are all similar. It is known (see Proposition 3) that a minimal WA can be obtained from a WA by alternating between two constructions which reduce the dimension to make it match the one of the span of the left (resp. right) reachability set. The result then follows from the next lemma, which states that both constructions decrease the length and dimension of the invariants. We prove it by considering an adequate change of basis, and verifying that it preserves invariants.

▶ **Lemma 26.** *Let $\mathcal{R}$ be a* WA *realizing a rational series f, let $S_{\mathcal{R}}$ be a Z-linear invariant of $\mathcal{R}$ of length n and dimension k and let $r = \dim(\mathsf{span}(\mathsf{LR}(\mathcal{R})))$. We can construct an r-dimensional* WA *$\mathcal{R}'$ realizing f, with a Z-linear invariant $S_{\mathcal{R}'}$ of length $\leq n$ and dimension $\leq k$. The same holds with $r = \dim(\mathsf{span}(\mathsf{RR}(\mathcal{R})))$.*

The next proposition allows to go from Z-linear invariants of WA to CRA. This construction builds on the one of [3, Lemma 3.13], in which they build an equivalent WA from the strongest Z-linear invariant of a WA. We show that an analogous construction is valid for any Z-linear invariant, and that we can use states of CRA to represent the different irreducible components of the invariant, thus reducing the number of registers used to the dimension of the invariant.

▶ **Proposition 27.** *Let $\mathcal{R}$ be a* WA*. If $\mathcal{R}$ has a Z-linear invariant of length n and dimension k, then there exists a linear* CRA *$\mathcal{A}$, with n states and k registers, such that $[\![\mathcal{A}]\!] = [\![\mathcal{R}]\!]$.*

The next proposition shows the converse direction, from CRA to invariants of WA. The construction is the classical one from CRA to WA. The existence of the adequate invariant follows from the determinism of the CRA which ensures that in any reachable configuration, only coordinates associated with the reachable state of the CRA can be non-zero.

▶ **Proposition 28.** *Let $\mathcal{A}$ be a linear* CRA*. If $\mathcal{A}$ has n states and k registers, then there exists a* WA *$\mathcal{R}$, with a Z-linear invariant of length n and dimension k, such that $[\![\mathcal{A}]\!] = [\![\mathcal{R}]\!]$.*

Using the three previous propositions, we can finally prove the main characterization:

Proof of Theorem 22. Given a linear CRA with n states and k registers, we can construct, thanks to Proposition 28, an equivalent WA with a Z-linear invariant of length n and dimension k. Then the desired minimal WA exists thanks to Proposition 25.

Reciprocally, applying the construction of Proposition 27 to any minimal WA gives the desired linear CRA. ◀

As we will discuss in the next subsection below, the three propositions we used for this proof can also be adapted to yield the same result for affine CRA.

4.4 Z-affine invariants and affine CRA

All the results of Section 4 can actually be extended to affine CRA using the *affine Zariski topology* instead of the linear one. It is a slight generalization of the linear Zariski topology where closed sets, called *Z-affine* sets, are finite unions of affine spaces instead of vector spaces, with lengths and dimensions defined like in the linear case. It is still a Noetherian topology coarser than the Zariski topology, affine maps are continuous and closed maps in this topology and, more broadly, it enjoys the same properties as the linear Zariski topology we considered throughout this section. For a set $S \subseteq \mathbb{K}^n$, we will denote by $\overline{S}^a$ it closure in the affine Zariski topology and, similarly to the linear case, for a WA $\mathcal{R} = (u, \mu, v)$, we will call any invariant of $\mathcal{R}$ that is a Z-affine set a *Z-affine invariant* of $\mathcal{R}$. Of course, the strongest Z-affine invariant of $\mathcal{R}$ is still the closure of its reachability set i.e. its "affine hull" $\overline{\mathsf{LR}(\mathcal{R})}^a$ and Remark 20 is still true for Z-affine invariants.

We obtain the same characterization of Theorem 22 in the affine setting :

▶ **Theorem 29** (Characterization). *Let f be a rational series. Then f can be realized by an affine* CRA *with n states and k registers iff there exists a minimal* WA *realizing f that has a Z-affine invariant of length at most n and dimension at most k.*

We can show that Propositions 25, 27 and 28 are also true if we replace Z-linear invariants by Z-affine ones and linear CRA by affine ones. So, the proof of Theorem 29 remains the same as Theorem 22. All the details can be found in Appendix A.2 of the full version of this paper [5].

Of course, this theorem has the same consequences of its linear counterpart and we obtain an affine version of Corollary 23

▶ **Corollary 30.** *The register complexity of a rational series f w.r.t. the class of affine* CRA *is the dimension of the strongest Z-affine invariant of any minimal* WA *realizing f.*

Working in the affine Zariski topology instead of the linear one can decrease the dimension of the strongest invariant by one, as shown in the following example.

▶ **Example 31.** On the alphabet $\Sigma = \{a\}$, let $\mathcal{R} = (u, \mu, v)$, where $u = (1, 2)$, $\mu(a) = \begin{pmatrix} 1 & 0 \\ 1 & 2 \end{pmatrix}$ and $v = (1, 0)^t$, be a WA (over $\mathbb{R}$) realizing the rational series f defined by $f(a^n) = \sum_{i=0}^{n} 2^i = 2^{n+1} - 1$.

The reachability set of $\mathcal{R}$ is $\mathsf{LR}(\mathcal{R}) = \left\{ \left(\sum_{i=0}^{n} 2^i, 2^{n+1}\right) \;\middle|\; n \in \mathbb{N} \right\}$.

For the linear Zariski topology, $\mathsf{LR}(\mathcal{R})$ is dense in $\mathbb{R}^2$. So the strongest Z-linear invariant $\overline{\mathsf{LR}(\mathcal{R})}^{\ell} = \mathbb{R}^2$ is two-dimensional. However, note that, for all $(x, y) \in \mathsf{LR}(\mathcal{R})$, $y = x + 1$. So, by an argument of density in the affine Zariski topology, the strongest Z-affine invariant $\overline{\mathsf{LR}(\mathcal{R})}^a$ is the affine line $y = x + 1$, which is one-dimensional.

Thus, in the case where the dimensions of the affine and linear hulls doesn't match, using affine CRA instead of linear CRA can allow to save one register :

▶ **Example 32** (Example 31 continued). The two CRA depicted on Figure 3 both realize the function of Example 31. On the left we have a linear CRA with two registers and, on the right, an affine CRA with only one register. The characterization theorems show that both have the minimal number of registers for their respective classes of CRA.

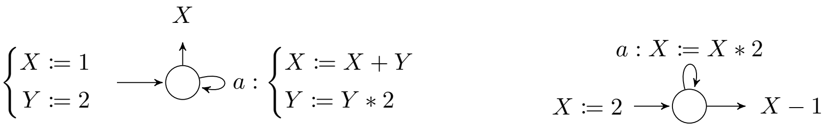

Figure 3 Two CRA detailed in Example 32.

5 Algorithms and complexity for the minimization problems

We present two original algorithms to solve the minimization problems we consider. It is worth observing the difference between the two characterizations we have obtained: while the register complexity can be computed from a canonical object (the strongest Z-linear invariant of the WA), the state-register complexity is based on the existence of a particular Z-linear invariant. This explains why we derive a non-deterministic procedure for the latter, and a deterministic for the former.

5.1 Algorithm for the state-register minimization problem

We provide here a NExpTime algorithm for the state-register minimization problem, hence proving Theorem 14. The algorithm runs in NPTime in n, k, and the size of the automaton. The fact that n is given in binary explains the exponential discrepancy.

Small representations of Z-affine sets. Let $\mathcal{R} = (u, \mu, v)$ be a WA of dimension d over an alphabet Σ. Let $L = A_1 \cup \cdots \cup A_n$ be a Z-affine set of length n of $\mathbb{K}^d$.

An $\mathcal{R}$-representation R of L is a set of n finite sets of words $S_1, \ldots, S_n$ such that $\mathsf{aff}\left(\{u\mu(w) |\ w \in S_i\}\right) = A_i$ for all $i \in \{1, \cdots, n\}$. The *size* of R is the sum of the lengths of all words appearing in R. The following key lemma shows that all Z-affine invariants of $\mathcal{R}$ have small $\mathcal{R}$-representations, up to considering stronger invariants.

▶ **Lemma 33.** *Let $\mathcal{R}$ be a* WA*. Let I be a Z-affine invariant of $\mathcal{R}$ of length n and dimension k. There exists an $\mathcal{R}$-representation R of size $\leq n^2k^2$ of a Z-affine invariant $J \subseteq I$, of dimension $\leq k$ and length $\leq n$.*

This property allows to derive the non-deterministic algorithm. First, minimization of a WA over a field can be performed in polynomial time (see e.g. [16, Corollary 4.17]). Then, let $\mathcal{R}$ be a minimal WA and let k, n be positive integers. From Lemma 33, we know that a Z-affine invariant of dimension k and length n can be represented in size $O(k^2n^2)$ (up to finding a stronger invariant with smaller dimension and length). The algorithm works thusly: first step is to guess an $\mathcal{R}$-representation R of a Z-affine set. The second step is to check that R represents an invariant, which can be done easily using basic linear algebra. From this one can compute an affine CRA with k registers and n states. Moreover, if we require that R is Z-linear, we obtain a linear CRA. If R is not an invariant, the computation rejects. Note that different accepting computations may give rise to different invariants and thus different CRAs.

5.2 Algorithm for the computation of Z-affine invariants

We describe a deterministic procedure which, given a WA $\mathcal{R}$ and an integer c, returns a Z-affine invariant J which is stronger that any Z-affine invariant I of $\mathcal{R}$ of length at most c. When c is chosen large enough, this procedure returns the strongest Z-affine invariant of $\mathcal{R}$. A similar procedure works as well for the computation of Z-linear invariants.

Algorithm 1 Computing a Z-affine invariant.

Require: A WA $\mathcal{R} = (u, \mu, v)$ of dimension d, an integer c
Ensure: A Z-affine invariant J of $\mathcal{R}$ stronger than $I_c(\mathcal{R})$

```
1: J := {u}
2: while J is not an invariant of R do
3:     Pick some component A of J, and some matrix M of R s.t. A · M ⊈ J
4:     J := J ∪ A · M
5:     if length(J) > c^d then
6:         J := REDUCE(J)
7:     end if
8: end while
9: return J
```

Intuitively, this procedure will build a Z-affine set J as follows: it starts with a set containing only the initial vector of $\mathcal{R}$, and incrementally extends it until it forms an invariant. During this process, it should ensure that J is included in every Z-affine invariant I of $\mathcal{R}$ of length at most c. This relies on the following easy observation: if such an invariant I contains at least $c+1$ points on the same affine line (*i.e.* a 1-dimensional affine space, denoted D), then I must have a component that contains D. Indeed, as I has length at most c, one of its components contains two such points. As this component is irreducible, it is an affine subspace, hence contains D. This reasoning can be lifted to higher dimensions as follows.

Given a WA $\mathcal{R}$, and $c \in \mathbb{N}$, we denote by $I_c(\mathcal{R}) = \bigcap_{\text{length}(I)\leq c} I$ the intersection of all Z-affine invariants of $\mathcal{R}$ with at most c components.

▶ **Lemma 34.** *Let $\mathcal{R}$ be a* WA *and let $c, k \in \mathbb{N}$. Let $A_1, \dots, A_{c^k+1} \subseteq I_c(\mathcal{R})$ be affine spaces such that: for any $P \subseteq [\![1, c^k+1]\!]$ with $|P| \geq c^{k-1}+1$,* $\mathsf{aff}(\cup_{i\in P} A_i)$ *has dimension k. Then* $\mathsf{aff}\left(\cup_{i\in[\![1,c^k+1]\!]} A_i\right) \subseteq I_c(\mathcal{R})$.

Using this lemma, we derive an effective procedure to simplify a Z-affine set $J = A_1 \cup \dots \cup A_{c^d+1}$ by "merging" two components. We denote by REDUCE(J) the resulting set.

▷ Claim 35. Let $\mathcal{R} = (u, \mu, v)$ be a WA of dimension d, let $c \in \mathbb{N}$. Let $A_1, \dots, A_{c^d+1} \subseteq I_c(\mathcal{R})$ be affine spaces. One can find $1 \leq i < j \leq c^d + 1$ such that $\mathsf{aff}(A_i \cup A_j) \subseteq I_c(\mathcal{R})$, in time $O(c^{p(d)})$, for some fixed polynomial p.

▶ **Theorem 36.** *Algorithm 1 is correct and terminates in time $O(c^{p(d)})$.*

Proof. Let us first discuss termination. Because of line 5-7, the length of J is at most $c^d + 1$. Moreover J is an increasing Z-affine set, thus its value can be modified at most $(d+1)\cdot(c^d+1)$ times, thus from Claim 35 the algorithm terminates in time $O(c^{p(d)})$.

We now discuss correctness. We need to show that J is stronger than $I_c(\mathcal{R})$. Initially, this holds. Moreover, if $A \subseteq I_c(\mathcal{R})$ is an affine set, then for any $M \in \mu(\Sigma)$, $A \cdot M \subseteq I_c(\mathcal{R})$, since $I_c(\mathcal{R})$ is invariant. Thus, line 4 preserves the property that J is stronger than $I_c(\mathcal{R})$. Using Claim 35, the REDUCE subroutine also preserves this property, since it only merges components whose affine span is contained in $I_c(\mathcal{R})$. ◀

5.3 Complexity of the register minimization problem

In order to compute the strongest Z-linear and Z-affine invariants of a WA using Algorithms 1, it is sufficient to be able to bound their lengths. The following result gives such bounds.

▶ **Theorem 37.** *Let $\mathcal{R} = (u, \mu, v)$ be a d-dimensional* WA *on a finite alphabet Σ. We have the following upper bounds :*

- *The lengths of $\overline{\mathsf{LR}(\mathcal{R})}^{\ell}$ and $\overline{\mathsf{LR}(\mathcal{R})}^{a}$ are at most doubly-exponential in d.*
- *If $\langle \mu(\Sigma) \rangle$ is commutative (*e.g. *Σ is unary), then the length of $\overline{\mathsf{LR}(\mathcal{R})}^{\ell}$ is at most exponential in d.*

We also have the following lower bound (which also hold for WA *over a unary alphabet):*

- *For all $d > 0$, there exist a d-dimensional* WA *having strongest Z-linear and Z-affine invariants with lengths exponential in d.*

Proof sketch. The first item is shown in [4], where the authors sketch a proof of a double-exponential upper bound on the length of the strongest Z-linear invariant of a WA, using tools from algebraic geometry, which holds for $\mathbb{Q}$ in particular and for any field $\mathbb{K}$ where there is a double-exponential bound on the maximal order of finite groups of invertible matrices (see [4, Proposition 48 and Remark 41]). Their proof can be adapted to $\overline{\mathsf{LR}(\mathcal{R})}^{a}$. The proof of the second item relies on basic linear algebra and on results and ideas from [4] for invertible matrices (see [4, Lemma 13 and Theorem 10]). Last, the lower bound is shown using a family of WA $(\mathcal{R}_i)_{i \in \mathbb{N}}$ whose dimension is polynomial in i and strongest Z-linear invariant has a length that is exponential in i. It is defined, using permutation matrices of dimension p, for some prime number p, which generate cyclic groups. The family is obtained by using block matrices composed of such permutation matrices. All the details are given in Appendix B.4 of the full version of this paper [5]. ◀

Thanks to this theorem, using Algorithm 1 with a large enough c (at most doubly-exponential in the dimension of the given WA), and thanks to Theorem 36, we can prove the following result:

▶ **Theorem 38.** *The strongest Z-linear/affine invariant of a* WA *is computable in 2-*EXPTIME*.*

This allows us to prove Theorem 12. Indeed, given a WA $\mathcal{R}$, we first compute an equivalent minimal WA, which can be done in polynomial time (see e.g. [16, Corollary 4.17]). Then, using Algorithm 1, we compute the strongest Z-linear (resp. Z-affine) invariant of $\mathcal{R}$. Corollary 23 (resp. Corollary 30) ensures that its dimension is the register complexity of f w.r.t. the class of linear (resp. affine) CRA, and the effectiveness follows from Proposition 27 (resp. its affine version).

Moreover, thanks to Theorem 38 and the results of [3], we also have:

▶ **Theorem 39.** *The sequentiality and unambiguity of a rational series are in 2-*EXPTIME*.*

Note that the complexities of the last two theorems drop down to EXPTIME when we have a simply exponential bound on the length of the strongest invariant. This is the case when one considers unary alphabets or WA with commuting transition matrices in the linear setting, as stated in Theorem 37. In these cases, the bound is sharp. It is still not clear however whether it is possible to close the gap between the bounds in the general case.

▶ Remark 40. It is also worth noting that, while the characterizations that we obtained are valid for any field, the complexities of the algorithms are given in terms of number of elementary operations over the considered field. Which means that they hold for fields where we can perform basic operations in polynomial time (such as $\mathbb{Q}$ or its finite extensions). Moreover, the general upper bounds on the lengths given by Theorem 37 were proven only for fields verifying a specific property (which is verified by $\mathbb{Q}$). See the proof for more details.

5.4 State/register tradeoff

Reducing the number of registers may increase the number of states and vice-versa. The following theorem summarizes what we know on this tradeoff.

▶ **Theorem 41.** *Let f be a rational series realized by some d-dimensional* WA $\mathcal{R}$*. Consider some pair of integers (n,k) optimal for f w.r.t. the class of linear* CRA*. The inequalities* $1 \leq n \leq \text{length}(\overline{\mathsf{LR}\,(\mathcal{R})}^{\ell}) = O(2^{2^d})$ *and* $\dim(\overline{\mathsf{LR}\,(\mathcal{R})}^{\ell}) \leq k \leq d$ *hold true.*

(They are valid in the affine setting as well)

▶ Remark 42. Building the CRA from the strongest invariant is not always optimal. There are some cases where it is possible to reduce the number of states of a CRA exponentially, while keeping the minimal number of registers, by choosing an invariant that is weaker than the strongest Z-linear/Z-affine invariant but shorter.

6 Conclusion

We have shown how to decide variants of CRA minimization problems, and have given complexity for the respective algorithms. There are several ways in which these algorithms could be improved. First, it would be worth reducing the gap between the lower and the upper bounds on the length of the strongest Z-linear invariant. Second, identifying a canonical invariant associated with the state-register minimization problem would allow to derive a deterministic algorithm for this problem. Third, one could hope for better complexity if one only considers the existence of equivalent CRA. For instance, in [13] the authors give a PSpace algorithm for the determinization problem (i.e. 1-register minimization problem) in the case of a polynomially ambiguous automaton, via a quite different approach.

Another line of research consists in trying to use the techniques we developed to solve the register minimization problem for other classes of CRA, for instance copyless CRA (which correspond to multi-sequential WA). Another ambitious goal is to consider register minimization in the context of different semirings, but there all the linear algebra tools which are crucial to solving these problems completely break down. Similarly, it seems that register minimization for polynomial automata would be very difficult: it was shown recently in [12] that the strongest algebraic invariant of a polynomial automaton is not computable. One possibility may be to bound the “degree” of the invariants, where Z-affine sets would correspond to algebraic sets of degree one.

References

1 Rajeev Alur, Loris D’Antoni, Jyotirmoy V. Deshmukh, Mukund Raghothaman, and Yifei Yuan. Regular functions and cost register automata. In *28th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2013, New Orleans, LA, USA, June 25-28, 2013*, pages 13–22. IEEE Computer Society, 2013.

2 Rajeev Alur and Mukund Raghothaman. Decision problems for additive regular functions. In *Automata, Languages, and Programming - 40th International Colloquium, ICALP 2013, Riga, Latvia, July 8-12, 2013, Proceedings, Part II*, volume 7966 of *Lecture Notes in Computer Science*, pages 37–48. Springer, 2013.

3 Jason Bell and Daniel Smertnig. Noncommutative rational pólya series. *Selecta Mathematica*, 27, July 2021. doi:10.1007/s00029-021-00629-2.

4 Jason P. Bell and Daniel Smertnig. Computing the linear hull: Deciding deterministic? and unambiguous? for weighted automata over fields. In *38th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2023, Boston, MA, USA, June 26-29, 2023*, pages 1–13. IEEE, 2023. doi:10.1109/LICS56636.2023.10175691.

5 Yahia Idriss Benalioua, Nathan Lhote, and Pierre-Alain Reynier. Minimizing cost register automata over a field, 2024. arXiv:2307.13505.

6 Jean Berstel. *Transductions and context-free languages*, volume 38 of *Teubner Studienbücher : Informatik*. Teubner, 1979. URL: https://www.worldcat.org/oclc/06364613.

7 Christian Choffrut. Une caractérisation des fonctions séquentielles et des fonctions sous-séquentielles en tant que relations rationnelles. *Theor. Comput. Sci.*, 5(3):325–337, 1977. doi:10.1016/0304-3975(77)90049-4.

8 Thomas Colcombet and Daniela Petrisan. Automata in the category of glued vector spaces. In Kim G. Larsen, Hans L. Bodlaender, and Jean-François Raskin, editors, *42nd International Symposium on Mathematical Foundations of Computer Science, MFCS 2017, August 21-25, 2017 - Aalborg, Denmark*, volume 83 of *LIPIcs*, pages 52:1–52:14. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2017. doi:10.4230/LIPIcs.MFCS.2017.52.

9 Laure Daviaud. Containment and equivalence of weighted automata: Probabilistic and max-plus cases. In Alberto Leporati, Carlos Martín-Vide, Dana Shapira, and Claudio Zandron, editors, *Language and Automata Theory and Applications - 14th International Conference, LATA 2020, Milan, Italy, March 4-6, 2020, Proceedings*, volume 12038 of *Lecture Notes in Computer Science*, pages 17–32. Springer, 2020. doi:10.1007/978-3-030-40608-0_2.

10 Laure Daviaud, Ismaël Jecker, Pierre-Alain Reynier, and Didier Villevalois. Degree of sequentiality of weighted automata. In *FOSSACS 2017*, volume 10203 of *Lecture Notes in Computer Science*, pages 215–230, 2017.

11 Laure Daviaud, Pierre-Alain Reynier, and Jean-Marc Talbot. A generalised twinning property for minimisation of cost register automata. In *LICS '16*, pages 857–866. ACM, 2016.

12 Ehud Hrushovski, Joël Ouaknine, Amaury Pouly, and James Worrell. On strongest algebraic program invariants. *J. ACM*, 70(5):29:1–29:22, 2023. doi:10.1145/3614319.

13 Ismaël Jecker, Filip Mazowiecki, and David Purser. Determinisation and unambiguisation of polynomially-ambiguous rational weighted automata. In Pawel Sobocinski, Ugo Dal Lago, and Javier Esparza, editors, *Proceedings of the 39th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2024, Tallinn, Estonia, July 8-11, 2024*, pages 46:1–46:13. ACM, 2024. doi:10.1145/3661814.3662073.

14 Heiko Vogler Manfred Droste, Werner Kuich. *Handbook of Weighted Automata*. Springer Berlin, Heidelberg, 2009. doi:10.1007/978-3-642-01492-5.

15 Mehryar Mohri. Finite-state transducers in language and speech processing. *Comput. Linguistics*, 23(2):269–311, 1997.

16 Jacques Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, 2009. doi:10.1017/CBO9781139195218.

17 Marcel Paul Schützenberger. On the definition of a family of automata. *Inf. Control.*, 4(2-3):245–270, 1961. doi:10.1016/S0019-9958(61)80020-X.