A Generalization of Semenov's Theorem to Automata over Real Numbers*

Bernard Boigelot[1], Julien Brusten[1]**, and Jérôme Leroux[2]

[1] Institut Montefiore, B28
Université de Liège
B-4000 Liège, Belgium
{boigelot,brusten}@montefiore.ulg.ac.be

[2] Laboratoire Bordelais de Recherche en Informatique (LaBRI)
351, cours de la Libération
F-33405 Talence Cedex, France
leroux@labri.fr

Abstract This work studies the properties of finite automata recognizing vectors with real components, encoded positionally in a given integer numeration base. Such automata are used, in particular, as symbolic data structures for representing sets definable in the first-order theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, i.e., the mixed additive arithmetic of integer and real variables. They also lead to a simple decision procedure for this arithmetic.

In previous work, it has been established that the sets definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$ can be handled by a restricted form of infinite-word automata, weak deterministic ones, regardless of the chosen numeration base. In this paper, we address the reciprocal property, proving that the sets of vectors that are simultaneously recognizable in all bases, by either weak deterministic or Muller automata, are those definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. This result can be seen as a generalization to the mixed integer and real domain of Semenov's theorem, which characterizes the sets of integer vectors recognizable by finite automata in multiple bases. It also extends to multidimensional vectors a similar property recently established for sets of numbers.

As an additional contribution, the techniques used for obtaining our main result lead to valuable insight into the internal structure of automata recognizing sets of vectors definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. This structure might be exploited in order to improve the efficiency of representation systems and decision procedures for this arithmetic.

---

* This work is supported by the *Interuniversity Attraction Poles* program *MoVES* of the Belgian Federal Science Policy Office, by the grant 2.4530.02 of the Belgian Fund for Scientific Research (F.R.S.-FNRS), and by the French project ANR-06-SETI-001 AVERISS.

** Research fellow ("Aspirant") of the Belgian Fund for Scientific Research (F.R.S.-FNRS).

1 Introduction

In the program analysis and verification field, one often faces the problem of finding a suitable formalism for expressing the constraints to be satisfied by the system configurations. Ideally, this formalism has to be decidable, while still remaining expressive enough for handling the class of constraints needed by the application. An example of such formalism is Presburger arithmetic, i.e., the first-order additive theory of integers $\langle \mathbb{Z}, +, \leq \rangle$, which is widely used for reasoning about programs manipulating integer variables. Presburger arithmetic is indeed decidable [1,2], yet expressive enough for describing arbitrary linear constraints as well as discrete periodicities [3].

A simple approach to deciding Presburger arithmetic consists in using finite automata. It is indeed known that, using the positional notation for encoding numbers and vectors into words, all Presburger-definable sets are mapped onto regular languages and can thus be recognized by automata [2,3]. A Presburger formula can be decided by recursively constructing an automaton recognizing its solutions, and then checking whether this automaton accepts a nonempty language. In some program verification applications, such automata, called *Number Decision Diagrams (NDDs)* are actually used as data structures for representing and manipulating symbolically the sets of program configurations that need to be handled [4].

Although every subset of $\mathbb{Z}^n$ that is Presburger-definable can be recognized by a finite automaton, the reciprocal property does not hold. For instance, denoting by $r \in \mathbb{N}_{>1}$ the base chosen for encoding numbers, the set $\{r^k \mid k \in \mathbb{N}\}$, which is not Presburger-definable, clearly corresponds to a regular language and is thus recognizable. The well-known Cobham's theorem states that, if a set $S \subseteq \mathbb{Z}$ is simultaneously recognizable by finite automata in two bases $r, s \in \mathbb{N}_{>1}$ that are multiplicatively independent, i.e., such that $r^p \neq s^q$ for all $p, q \in \mathbb{N}_{>0}$, then S is Presburger definable [5]. This result has then been extended to subsets of $\mathbb{Z}^n$, with $n > 0$, i.e., sets of integer vectors, by Semenov [6]. As a corollary of Semenov's theorem, the subsets of $\mathbb{Z}^n$ that are recognizable by finite automata in every base $r \in \mathbb{N}_{>1}$ are exactly those that are Presburger-definable.

Quite recently, automata recognizing sets of numbers and vectors have been generalized to the mixed integer and real domain [7]. In this setting, the base-r encoding of numbers and vectors take the form of infinite words over the alphabet $\{0, 1, \ldots, r-1, \star\}$, where "$\star$" is a separator symbol used for distinguishing their integer and fractional parts. A *Real Vector Automaton (RVA)* recognizing a set $S \subseteq \mathbb{R}^n$ is then an infinite-word automaton that accepts the encodings of the elements of S.

It is worth stressing out that RVA are not only theoretical objects; they are used as actual data structures in verification tools such as LASH for representing symbolically the sets of configurations of programs relying on both integer and real variables during their state-space exploration [8]. The decision tool LIRA also uses RVA for representing the set of solutions of mixed real and integer arithmetic formulas [9]. For such applications, it is not sufficient to establish that all sets of interest are representable by RVA and that all the needed operations are computable on them, but also to obtain a symbolic representation system that is concise enough for handling complex sets using a reasonable amount of memory, and for which the manipulation algorithms are efficient. In particular, this precludes the use of unrestricted infinite-word automata for describing RVA, due to the difficulty of carrying out some operations such as set complementation [13]. It is therefore essential to define restricted forms of RVA that can be efficiently handled, and to precisely characterize their expressiveness in order to match the requirements of the intended applications. Another goal is to investigate whether the transition relation of these restricted RVA has structural properties that can be exploited in order to represent them more efficiently.

In previous work, a result analogous to Cobham's theorem has been obtained for RVA: The sets $S \subseteq \mathbb{R}$ that are recognizable by RVA in two bases that do not share the same set of prime factors[1] are exactly those that are definable in the first-order theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, i.e., the extension of Presburger arithmetic to mixed integer and real variables [10]. This has an important consequence. One indeed knows that the full expressive power of ω-regular languages is not needed for representing the sets definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, since those sets can be recognized by *weak deterministic automata* [11], a restricted class of infinite-word automata that are much more easily manipulated algorithmically, and admit a canonical form. It follows that the sets of reals that are recognizable by RVA in every base $r \in \mathbb{N}_{>1}$ are exactly those that can be recognized by weak deterministic RVA.

This paper is aimed at extending this result to sets of vectors with a fixed dimension, i.e., to subsets of $\mathbb{R}^n$ with $n > 0$. This can be seen as a generalization of Semenov's theorem to real vectors. Precisely, we prove that the sets $S \subseteq \mathbb{R}^n$ that are simultaneously recognizable by RVA in two bases that do not share the same set of prime factors are those that are definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. From this result, it follows that the sets of vectors that are recognizable by RVA in every base $r \in \mathbb{N}_{>1}$ are exactly those that can be recognized by weak deterministic RVA. The same proof also establishes that the sets that are recognizable by weak deterministic RVA in two multiplicatively independent bases are definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$ as well. Those results are significant for practical applications, since they imply that weak deterministic automata can be used for implementing RVA in all cases where the sets of vectors that are symbolically represented are expressed as combinations of linear constraints and discrete periodicities.

As an additional contribution, the techniques used for obtaining this result give out valuable insight into the internal structure of RVA recognizing sets of vectors definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. It might be possible to exploit this structure in order to improve the efficiency of symbolic representation systems and decision procedures for that arithmetic.

[1] As opposed to the integer case, it has been shown that the result does not hold for multiplicatively independent bases [10].

2 Preliminaries

2.1 Positional Encoding of Vectors

Let $r \in \mathbb{N}_{>1}$ be a *base*. The positional notation in base r *encodes* numbers $x \in \mathbb{R}$ as infinite words of the form $w_I \star w_F$ over the finite alphabet $\Sigma_r \cup \{\star\}$, where $\Sigma_r = \{0, 1, \ldots, r-1\}$, and "$\star$" is a separator symbol. The finite prefix $w_I \in \Sigma_r^+$ and the infinite suffix $w_F \in \Sigma_r^\omega$ respectively encode an *integer* and a *fractional part* of x. In other words, w_I encodes a number $x_I \in \mathbb{Z}$, and w_F a number $x_F \in [0, 1]$, such that $x_I + x_F = x$. Note that integer numbers admit two decompositions into integer and fractional parts, e.g., $x = 2$ leads to both $x_I = 2$ and $x_F = 0$, and $x_I = 1$ and $x_F = 1$.

An encoding w_I of an integer part $x_I \in \mathbb{N}$ is a word $a_{p-1}a_{p-2} \ldots a_0 \in \Sigma_r^+$, with $p > 0$, such that $x_I = \sum_{i=0}^{p-1} a_i r^i$. For signed numbers $x_I \in \mathbb{Z}$, the *r's-complement* representation is used, implying that the *sign* digit a_{p-1} is then equal to 0 for positive (or zero) numbers, and to $r-1$ for negative ones. For a negative number, the value of x_I becomes equal to $-r^p + \sum_{i=0}^{p-1} a_i r^i$. The length p of encodings is not fixed, but chosen large enough for satisfying the constraint $-r^{p-1} \leq x_I < r^{p-1}$. Finally, an encoding w_F of a fractional part $x_F \in [0, 1]$ is a word $b_0 b_1 \ldots \in \Sigma_r^\omega$ such that $x_F = \sum_{i \geq 0} b_i r^{-i}$.

This encoding scheme can be extended to vectors $\boldsymbol{x} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, with $n > 0$. The idea is to encode each component x_i separately into a word w_i, but in such a way that these words share the same integer-part length. This can always be achieved, for the sign digit of an encoding can be repeated at will without altering the encoded value. One thus obtains a vector $(w_1, w_2, \ldots, w_n)$ of encodings in which the separator symbol "$\star$" occurs at the same position in each component. By reading those components synchronously one symbol at a time, one eventually obtains an encoding of $\boldsymbol{x}$ as a single word $w_I \star w_F$ over the n-dimensional alphabet Σ_r^n, augmented with a unique separator symbol "$\star$". For each word $w \in \{0, r-1\}^n (\Sigma_r^n)^* \star (\Sigma_r^n)^\omega$, the vector $\boldsymbol{x} \in \mathbb{R}^n$ encoded by w in base r is denoted by $[w]_r$.

2.2 Real Vector Automata

Consider a base $r \in \mathbb{N}_{>1}$ and a set $S \subseteq \mathbb{R}^n$, with $n > 0$. Let $L(S) \subseteq (\Sigma_r^n)^+ \star (\Sigma_r^n)^\omega$ denote the language formed by the encodings of the elements of S. If $L(S)$ is ω-regular, then any infinite-word automaton that accepts $L(S)$ is a *Real Vector Automaton (RVA)* recognizing S [7]. In this paper, in order to simplify some developments thanks to their deterministic transition relation, we assume w.l.o.g. that RVA take the form of *Muller automata* [12]. A set $S \subseteq \mathbb{R}^n$ that can be recognized by a RVA in base r is said to be *r-recognizable*.

For practical applications as symbolic representations of sets, infinite-word automata are somehow problematic, since some of their manipulation algorithms are known to be significantly costlier than their finite-word counterparts [13]. In the case of RVA, its has been shown that the full expressive power of infinite-word automata is not needed for recognizing the sets definable in the first-order theory

$\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. Indeed, for any base $r \in \mathbb{N}_{>1}$, such sets can be recognized by *weak deterministic RVA* [11]. A weak RVA representing a set S is a Büchi automaton accepting $L(S)$, such that each of its strongly connected components is either globally accepting or globally non-accepting. Compared to general infinite-word automata, weak deterministic ones are much more easily manipulated algorithmically. In addition, they admit a canonical form that simplifies comparison operations between symbolically represented sets [14]. We will say that a set $S \subseteq \mathbb{R}^n$ that can be recognized by a weak deterministic RVA in base r is *weakly r-recognizable*.

2.3 Properties of Recognizable Sets

In the next sections, we study the properties of sets that are recognizable, or weakly recognizable, in one or several bases. Such sets are characterized by the following results.

Theorem 1 ([15]). *Let $n \in \mathbb{N}_{>0}$ and $r \in \mathbb{N}_{>1}$. A set $S \subseteq \mathbb{R}^n$ is r-recognizable iff it is definable in the first-order theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_r \rangle$, where $X_r \subset \mathbb{R}^3$ is a base-dependent predicate such that $X_r(x, u, k)$ holds whenever u is an integer power of r, and there exists an encoding of x in which the digit at the position specified by u is equal to k.*

Theorem 2 ([11]). *Let $n \in \mathbb{N}_{>0}$ and $r \in \mathbb{N}_{>1}$. A set $S \subseteq \mathbb{R}^n$ is weakly r-recognizable iff it is r-recognizable, and it belongs to the topological class $F_\sigma \cap G_\delta$ of the metric topology over $\mathbb{R}^n$ induced by the Euclidean distance. This means that the set has to be decomposable both into a countable union of closed sets, and into a countable intersection of open sets.*

In particular, it is known that every subset of $\mathbb{R}^n$ that is definable in the first-order theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, i.e., the extension of Presburger arithmetic to mixed integer and real variables, satisfies the hypotheses of Theorem 2, and it is therefore weakly recognizable in every base $r \in \mathbb{N}_{>1}$ [11].

The following theorems and lemmas introduce some operations and transformations that preserve the recognizable nature of sets.

Theorem 3. *Let $n \in \mathbb{N}_{>0}$ be a dimension, and $r, s \in \mathbb{N}_{>1}$ be* multiplicatively dependent *bases, i.e., such that $r^k = s^l$ for some $k, l \in \mathbb{N}_{>0}$. A set $S \subseteq \mathbb{R}^n$ is (resp. weakly) r-recognizable iff it is (resp. weakly) s-recognizable.*

Proof sketch. From Theorems 1 and 2, it suffices to show that definability in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_r \rangle$ is equivalent to definability in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_s \rangle$. Furthermore, since $r^k = s^l$, it actually suffices to establish that definability in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_t \rangle$ is equivalent to definability in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_{t^i} \rangle$ for all $t \in \mathbb{N}_{>1}$ and $i \in \mathbb{N}_{>0}$. An encoding of a number $x \in \mathbb{R}$ in base t^i can directly be turned into one of the same number in base t by replacing each digit (belonging to the alphabet Σ_{t^i}) into a sequence of i digits from Σ_t. The reciprocal operation is similar. It follows that the predicate X_t can be defined in terms of X_{t^i}, and reciprocally. □

Theorem 3 states that (resp. weak) recognizability in bases that are multiplicatively dependent is equivalent to (resp. weak) recognizability in one of them. In the sequel, we will thus only consider bases r and s that are multiplicatively independent.

Lemma 1. *Let $n \in \mathbb{N}_{>0}$, $S_1, S_2 \subseteq \mathbb{R}^n$, and $r \in \mathbb{N}_{>1}$. If S_1 and S_2 are both (resp. weakly) r-recognizable, then the sets $S_1 \cup S_2$, $S_1 \cap S_2$, $S_1 \setminus S_2$ and $S_1 \times S_2$ are (resp. weakly) r-recognizable as well.*

Proof sketch. The class of sets definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_r \rangle$ is closed under Boolean and Cartesian operators. The same property holds for the topological class $F_\sigma \cap G_\delta$. The result then follows from Theorems 1 and 2. □

Lemma 2. *Let $n \in \mathbb{N}_{>0}$, $r \in \mathbb{N}_{>1}$, $C \in \mathbb{Q}^{n \times n}$ such that $det(C) \neq 0$, and $\boldsymbol{a} \in \mathbb{Q}^n$. If a set $S \subseteq \mathbb{R}^n$ is (resp. weakly) r-recognizable, then the set $CS + \boldsymbol{a}$ is (resp. weakly) r-recognizable as well.*

Proof sketch. The proof is by similar arguments as in that of Lemma 1. Indeed, the transformation $\boldsymbol{x} \mapsto C\boldsymbol{x} + \boldsymbol{a}$ is definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq, X_r \rangle$, and preserves the topological class $F_\sigma \cap G_\delta$. □

It is worth mentioning that, in the statement of Lemma 2, it is essential to require $\det(C) \neq 0$ as far as weak recognizability is concerned. Indeed, a transformation $\boldsymbol{x} \mapsto C\boldsymbol{x} + \boldsymbol{a}$ with a singular matrix C amounts to a *projection*, which generally alters topological properties of sets. As an example, the set $S = \{(zr^k, r^k) \mid k, z \in \mathbb{Z}\}$ belongs to $F_\sigma \cap G_\delta$, and it is actually weakly r-recognizable, whereas the set CS, with $C = \mathrm{diag}(1, 0)$, does not.

Although projection does not preserve weak recognizability, one can however sometimes extract from a set a weak recognizable set of smaller dimension. This operation is described by the following lemma.

Lemma 3. *Let $n, m \in \mathbb{N}_{>0}$, $r \in \mathbb{N}_{>1}$. If two sets $S_1 \subseteq \mathbb{R}^n$ and $S_2 \subseteq \mathbb{R}^m$ are such that $S_1 \times S_2$ is weakly r-recognizable, then S_1 and S_2 are weakly r-recognizable as well.*

Proof sketch. The proof is by similar arguments as in that of Lemma 1. □

Finally, the following result addresses the comparison of two recognizable sets.

Theorem 4. *Let $n \in \mathbb{N}_{>0}$ and $r \in \mathbb{N}_{>1}$. Two r-recognizable sets $S_1, S_2 \subseteq \mathbb{R}^n$ are equal iff they coincide over the rational vectors, i.e., iff $S_1 \cap \mathbb{Q}^n = S_2 \cap \mathbb{Q}^n$.*

Proof sketch. The vectors in $\mathbb{Q}^n$ are exactly those that are encoded by ultimately periodic words, i.e., words of the form uv^ω with $|v| \geq 1$. Two ω-regular languages are equal iff they coincide over the ultimately periodic words [16]. □

As a corollary, one can always extract a rational vector from a non-empty r-recognizable set.

3 Problem Reductions

Our main goal will be to prove that a set $S \subseteq \mathbb{R}^n$ that is recognizable or weakly recognizable in two bases r and s that are multiplicatively independent, with some possible additional restrictions on r and s, is necessarily definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. In this section, we reduce this general problem to simpler ones.

3.1 Reduction to $[0,1]^n$

It is known that a set $S \subseteq \mathbb{R}^n$ that is r-recognizable can be decomposed into a finite union $S = \bigcup_i (S_i^I + S_i^F)$, where the sets $S_i^I \subseteq \mathbb{Z}^n$ are non-empty and pairwise disjoint, and the sets $S_i^F \subseteq [0,1]^n$ are non-empty and pairwise different[2] [17,10]. This decomposition of S into sets S_i^I and S_i^F is independent from the base r. In addition, each set S_i^I is recognizable by a finite-word automaton in base r (operating only on the integer part of r-encodings), and each set S_i^F is (resp. weakly) r-recognizable if S is (resp. weakly) r-recognizable as well [17,10].

Consider two multiplicatively independent bases r and s, and a set $S \subseteq \mathbb{R}^n$ that is both r- and s-recognizable. Applying Semenov's theorem, one obtains that the sets S_i^I are definable in $\langle \mathbb{Z}, +, \leq \rangle$. It follows that, in order to prove that S is definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, it suffices to show that each set S_i^F is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. Since this theory is closed under elimination of quantifiers [18], this is equivalent to proving that each S_i^F can be expressed as a Boolean combination of linear constraints with rational coefficients.

We have therefore reduced our main problem from $\mathbb{R}^n$ to the simpler domain $[0,1]^n$. From now on, we will stay within $[0,1]^n$ and consider that RVA only recognize the fractional part of encodings, their integer part being restricted to zero. Formally, we introduce $[w]_r$, with $w \in (\Sigma_r^n)^\omega$, as a shorthand for $[\mathbf{0} \star w]_r$.

3.2 Reduction to Product-Stable Sets

In order to be able to prove that the recognizability of a subset of $[0,1]^n$ in multiple bases leads to its definability in $\langle \mathbb{R}, +, \leq, 1 \rangle$, we need to establish a link between the arithmetical properties of this set, and the structure of automata recognizing it.

Let $n \in \mathbb{N}_{>0}$, $r \in \mathbb{N}_{>1}$, and let $S \subseteq [0,1]^n$ be a set recognized in base r by a RVA $\mathcal{A}$. We associate to each state q of $\mathcal{A}$ the language $L(q)$ accepted from q, as well as the set of vectors $S(q) \subseteq [0,1]^n$ encoded by $L(q)$, i.e., $S(q) = \{[w]_r \mid w \in L(q)\}$.

Recall that $\mathcal{A}$ is a (deterministic) Muller automaton. For each finite path $q \xrightarrow{\sigma} q'$ of $\mathcal{A}$, the language $L(q')$ can be expressed as $L(q') = \sigma^{-1}L(q) = \{w \in (\Sigma_r^n)^\omega \mid \sigma w \in L(q)\}$. Similarly, the set $S(q')$ can be expressed in terms of $S(q)$. Denoting by $[\sigma]_r$ the integer vector encoded by σ, i.e., $[\sigma]_r = [\mathbf{0}\sigma \star \mathbf{0}^\omega]_r$, we get

$$S(q') = \left\{ \boldsymbol{x} \in [0,1]^n \mid \frac{[\sigma]_r + \boldsymbol{x}}{r^{|\sigma|}} \in S(q) \right\}.$$

[2] The property is actually known for the domain $\mathbb{R}$, but straightforwardly generalizes to $\mathbb{R}^n$.

From this relation and Lemmas 1 and 2, one obtains that $S(q')$ is (resp. weakly) recognizable in all bases for which $S(q)$ is (resp. weakly) recognizable.

Consider now the particular case $q = q'$, i.e., assume that the path labeled by σ cycles from q to itself. The previous relation becomes

$$\boldsymbol{x} \in S(q) \Leftrightarrow \boldsymbol{x} \in [0,1]^n \wedge \frac{[\sigma]_r + \boldsymbol{x}}{r^{|\sigma|}} \in S(q).$$

Remark that the transformation $\boldsymbol{x} \mapsto ([\sigma]_r + \boldsymbol{x})/r^{|\sigma|}$ admits the fixed point $\boldsymbol{x} = [\sigma]_r/(r^{|\sigma|} - 1) = [\sigma^\omega]_r \in [0,1]^n$. Translating $S(q)$ so as to move this fixed point onto $\mathbf{0}$, one gets

$$\boldsymbol{x} \in S(q) - [\sigma^\omega]_r \Leftrightarrow \boldsymbol{x} \in [0,1]^n - [\sigma^\omega]_r \wedge \frac{\boldsymbol{x}}{r^{|\sigma|}} \in S(q) - [\sigma^\omega]_r.$$

This prompts the following definition, adapted from [10].

Definition 1. *Let* $n \in \mathbb{N}_{>0}$*,* $\boldsymbol{v} \in [0,1]^n \cap \mathbb{Q}^n$ *be a* pivot*, and* $f \in \mathbb{R}_{\geq 1}$ *be a* factor*. A set* $S \subseteq [0,1]^n$ *is* f-product-stable *with respect to the pivot* $\boldsymbol{v}$ *iff* $\forall \boldsymbol{x} \in [0,1]^n - \boldsymbol{v} : \boldsymbol{x} \in S - \boldsymbol{v} \Leftrightarrow (1/f)\boldsymbol{x} \in S - \boldsymbol{v}$.

Intuitively, that a set is f-product-stable with respect to the pivot $\boldsymbol{v}$ means that the set does not change when it magnified by the zoom factor f around the fixed point $\boldsymbol{v}$. Remark that this property is preserved by transformations of the form $\boldsymbol{x} \mapsto C\boldsymbol{x} + \boldsymbol{a}$, with $C \in \mathbb{Q}^{n \times n}$ and $\boldsymbol{a} \in \mathbb{Q}^n$, provided that $[0,1]^n \subseteq C[0,1]^n + \boldsymbol{a}$, and the new pivot $\boldsymbol{v'} = C\boldsymbol{v} + \boldsymbol{a}$ belongs to $[0,1]^n$.

In summary, if $\mathcal{A}$ recognizes the set $S \subseteq [0,1]^n$ in base r, then each reachable state q of $\mathcal{A}$ recognizes a set $S(q) \subseteq [0,1]^n$ that is (resp.weakly) recognizable in all bases for which S is (resp. weakly) recognizable. Furthermore, if there exists a cycle $q \xrightarrow{\sigma} q$, then the set $S(q)$ is $r^{|\sigma|}$-product-stable with respect to the pivot $[\sigma^\omega]_r$. We have thus established a link between a structural property of $\mathcal{A}$ (the presence of a cycle rooted at q) and an arithmetical property of $S(q)$ (its product stability).

The next step is to show that any recognizable set can be decomposed into a combination of product-stable sets that can be considered individually.

Consider the set Q_1 of states q of $\mathcal{A}$ from which there exists a cycle $q \xrightarrow{\sigma} q$, with $\sigma \in (\Sigma_r^n)^+$. Note that every infinite path of $\mathcal{A}$ eventually visits a state in Q_1. Let L be the language of words $\sigma_i \in (\Sigma_r^n)^*$ labeling finite paths $\pi = q_0 \xrightarrow{\sigma_i} q_i$ such that q_0 is the initial state of $\mathcal{A}$, $q_i \in Q_1$, $q' \notin Q_1$ for every state q' distinct from q_i visited by π, and there is only one occurrence of q_i in π. The language L is finite, and it maps each $\sigma_i \in L$ to a state q_i of $\mathcal{A}$. For each such q_i, $\mathcal{A}$ admits a cycle rooted at q_i, hence there exists $\boldsymbol{v_i} \in \mathbb{Q}^n \cap [0,1]^n$ and $l_i \in \mathbb{N}_{>0}$ such that $S(q_i)$ is r^{l_i}-product-stable with respect to the pivot $\boldsymbol{v_i}$. Note that each $S(q_i)$ is (resp. weakly) recognizable in all bases for which S is (resp. weakly) recognizable. Moreover, since $S = \bigcup_{\sigma_i \in L} (1/r^{|\sigma_i|})(S(q_i) + [\sigma_i]_r)$, we have that S is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$ if all the sets $S(q_i)$ are definable in the same theory.

It follows from this result that, in order to prove that the recognizability of a set $S \subseteq [0,1]^n$ in multiple bases implies its definability in $\langle \mathbb{R}, +, \leq, 1 \rangle$, it is sufficient to prove this property for sets S that are r^l-product-stable, with $l \in \mathbb{N}_{>0}$.

4 Recognizability in Multiple Bases

In this section, we prove the following results, which generalize to $\mathbb{R}^n$ those developed in [17,10].

Theorem 5. *Let $n \in \mathbb{N}_{>0}$ be a dimension, and $r, s \in \mathbb{N}_{>1}$ be bases with different sets of prime factors (i.e., such that there exists a prime factor of one that does not divide the other). If a set $S \subseteq [0,1]^n$ is both r- and s-recognizable, then it is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$.*

Theorem 6. *Let $n \in \mathbb{N}_{>0}$ be a dimension, and $r, s \in \mathbb{N}_{>1}$ be multiplicatively independent bases. If a set $S \subseteq [0,1]^n$ is both weakly r- and weakly s-recognizable, then it is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$.*

Our approach is by induction on n. The case $n = 1$ is an immediate consequence of [17,10]. It remains to address the inductive case where $n \geq 2$, assuming that Theorems 5 and 6 hold for smaller dimensions. Exploiting the results of Section 3.2, we only consider w.l.o.g. sets $S \subseteq [0,1]^n$ that are r^l-product-stable for some $l \in \mathbb{N}_{>0}$ and pivot $\boldsymbol{v} \in [0,1]^n \cap \mathbb{Q}^n$.

4.1 Using s-recognizability

Consider a set $S \subseteq [0,1]^n$ that is recognizable in two bases $r, s \in \mathbb{N}_{>1}$. Assume that there exist $l \in \mathbb{N}_{>0}$ and $\boldsymbol{v} \in [0,1]^n \cap \mathbb{Q}^n$ such that S is r^l-product-stable with respect to the pivot $\boldsymbol{v}$. We show in this section that there exists an integer $l' \in \mathbb{N}_{>0}$ such that S is $s^{l'}$-product-stable as well.

We first consider the case $\boldsymbol{v} = \mathbf{0}$. Let $\mathcal{A}_s$ be a RVA recognizing S in base s. We assume w.l.o.g. that $\mathcal{A}_s$ has a complete transition relation, hence it admits an ultimately cyclic path labeled by $\mathbf{0}^\omega$, which we denote

$$q_0 \xrightarrow{\mathbf{0}^m} [q \xrightarrow{\mathbf{0}^{l'}}]^\omega,$$

where q_0 is the initial state of $\mathcal{A}_s$, $m \in \mathbb{N}$ and $l' \in \mathbb{N}_{>0}$. By the same reasoning as in Section 3.2, we obtain that the set $S(q)$ encoded in base s by the language $L(q)$ accepted from q in $\mathcal{A}_s$ is both r- and s-recognizable. Moreover, $S(q)$ is $s^{l'}$-product-stable with respect to the pivot $\mathbf{0}$.

Since $\mathcal{A}_s$ is deterministic, it admits only one path from q_0 to q labeled by $\mathbf{0}^m$, which leads to $S(q) = s^m S \cap [0,1]^n$. From this relation, and the r^l-product-stability hypothesis on S, it follows that $S(q)$ is r^l-product-stable as well, with respect to the pivot $\mathbf{0}$.

The set $S(q)$ is thus both r^l- and $s^{l'}$-product-stable, with respect to the same pivot $\mathbf{0}$. Let us show that these properties imply that S itself is both r^l- and $s^{l'}$-product-stable. By hypothesis, S is r^l-product-stable with respect to $\mathbf{0}$. For any $\boldsymbol{x} \in [0,1]^n$ and $k \in \mathbb{N}$, we thus obtain

$$\begin{aligned} \boldsymbol{x} \in S \;\Leftrightarrow\; \frac{1}{r^{lk}}\boldsymbol{x} \in S \;&\Leftrightarrow\; \frac{s^{|\sigma'|}}{r^{lk}}\boldsymbol{x} \in s^{|\sigma'|}S \\ &\Leftrightarrow\; \frac{s^{|\sigma'|}}{r^{lk}}\boldsymbol{x} \in S(q), \end{aligned}$$

if k is chosen large enough to have $r^{lk} \geq s^{|\sigma'|}$. Since $S(q)$ is $s^{l'}$-product-stable with respect to $\mathbf{0}$, we get

$$\begin{aligned} \frac{s^{|\sigma'|}}{r^{lk}}\boldsymbol{x} \in S(q) \;&\Leftrightarrow\; \frac{s^{|\sigma'|-l'}}{r^{lk}}\boldsymbol{x} \in S(q) \\ &\Leftrightarrow\; \frac{1}{r^{lk}s^{l'}}\boldsymbol{x} \in S \;\Leftrightarrow\; \frac{1}{s^{l'}}\boldsymbol{x} \in S. \end{aligned}$$

proving that S is $s^{l'}$-product-stable with respect to $\boldsymbol{v}$ in the special case $\boldsymbol{v} = \mathbf{0}$.

The general case $\boldsymbol{v} \in [0,1]^n \cap \mathbb{Q}^n$ is obtained by decomposing S according to the 2^n possible positions of vectors in $[0,1]^n$ with respect to $\boldsymbol{v}$. For each vector $\boldsymbol{a} \in \{-1,1\}^n$ we introduce the matrix $M_{\boldsymbol{a}} = \mathrm{diag}(\boldsymbol{a})$, the set $D_{\boldsymbol{a}} = (\boldsymbol{v} + M_{\boldsymbol{a}}\mathbb{R}^n_{\geq 0}) \cap [0,1]^n$ and the set $S_{\boldsymbol{a}} = S \cap D_{\boldsymbol{a}}$. Each set $D_{\boldsymbol{a}}$ is a Cartesian product of intervals $D_{\boldsymbol{a}} = I_1 \times \cdots \times I_n$, where for all $i \in \{1,\ldots,n\}$,

$$I_i = \begin{cases} [v_i, 1] \text{ if } a_i = 1, \\ [0, v_i] \text{ if } a_i = -1. \end{cases}$$

The vectors $\boldsymbol{a}$ such that $D_{\boldsymbol{a}}$ has a positive volume are identified by introducing the set A of vectors $\boldsymbol{a}$ such that $f_i(v_i) > 0$ for any $i \in \{1,\ldots,n\}$, where $f_i(v_i)$ denotes the length of the interval I_i, i.e

$$f_i(v_i) = \begin{cases} 1 - v_i \text{ if } a_i = 1, \\ \quad v_i \quad \text{if } a_i = -1. \end{cases}$$

Since each zero volume set is included into at least one positive volume set, we have $\bigcup_{\boldsymbol{a}\in A} D_{\boldsymbol{a}} = [0,1]^n$, which implies $S = \bigcup_{\boldsymbol{a}\in A} S_{\boldsymbol{a}}$.

For each $\boldsymbol{a} \in A$, the set $M_{\boldsymbol{a}}(D_{\boldsymbol{a}} - \boldsymbol{v})$ takes the form of the Cartesian product $[0, f_1(v_1)] \times \cdots \times [0, f_n(v_n)]$. We can thus map the elements of $D_{\boldsymbol{a}}$ onto $[0,1]^n$ by defining the transformation $\boldsymbol{x} \mapsto C_{\boldsymbol{a}}M_{\boldsymbol{a}}(\boldsymbol{x} - \boldsymbol{v})$, where $C_{\boldsymbol{a}} = \mathrm{diag}(1/f_1(v_1), \ldots, 1/f_n(v_n))$.

We now consider, for each $\boldsymbol{a} \in A$, the set $S'_{\boldsymbol{a}} = C_{\boldsymbol{a}}M_{\boldsymbol{a}}(S_{\boldsymbol{a}} - \boldsymbol{v})$. From the r^l-product-stability of S with respect to $\boldsymbol{v}$, it follows that $S'_{\boldsymbol{a}}$ is r^l-product-stable with respect to $\mathbf{0}$. By Lemmas 1 and 2, $S'_{\boldsymbol{a}}$ inherits the recognizability properties of S. Moreover, there exists $l'_{\boldsymbol{a}} \in \mathbb{N}_{>0}$ such that $S'_{\boldsymbol{a}}$ is $s^{l'_{\boldsymbol{a}}}$-product-stable with respect to $\mathbf{0}$. From this property and the equality $S_{\boldsymbol{a}} = \boldsymbol{v} + M_a^{-1}C_{\boldsymbol{a}}^{-1}S_{\boldsymbol{a}}$, we deduce that $S_{\boldsymbol{a}}$ is $s^{l'_{\boldsymbol{a}}}$-product-stable with respect to $\boldsymbol{v}$. From $S = \bigcup_{\boldsymbol{a}\in A} S_{\boldsymbol{a}}$, it then follows that S is $s^{l'}$-product-stable, where $l' = \mathrm{lcm}_{\boldsymbol{a}\in A}(l'_{\boldsymbol{a}})$.

In summary, we have established that S, in addition to being r^l-product-stable by hypothesis, is $s^{l'}$-product-stable as well. It remains to show that these properties, combined with our inductive hypotheses, imply that S is definable in $\langle \mathbb{R}, +, \leq, 1\rangle$.

4.2 Exploiting Multiple Product Stabilities

We thus consider a set $S \subseteq [0,1]^n$ and two bases $r, s \in \mathbb{N}_{>1}$, such that either r and s are multiplicatively independent and S is weakly r- and weakly s-recognizable, or r and s do not share the same set of prime factors[3] and S is r- and s-recognizable. Using the results of Section 4.1, we assume that there exist $l, l' \in \mathbb{N}_{>0}$ and $\boldsymbol{v} \in [0,1]^n \cap \mathbb{Q}^n$ such that S is both r^l- and $s^{l'}$-product-stable with respect to the pivot $\boldsymbol{v}$. Our goal is to show that S is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$.

We first prove the following property.

Property 1. For each $\boldsymbol{x} \in [0,1]^n \cap \mathbb{Q}^n$ such that $\boldsymbol{x} \neq \boldsymbol{v}$, let $h_{\boldsymbol{v}}(\boldsymbol{x})$ denote the set $\{\boldsymbol{v} + \lambda(\boldsymbol{x} - \boldsymbol{v}) \in [0,1]^n \mid \lambda \in \mathbb{R}_{>0}\}$. We have either $h_{\boldsymbol{v}}(\boldsymbol{x}) \subseteq S$, or $h_{\boldsymbol{v}}(\boldsymbol{x}) \cap S = \emptyset$.

Proof sketch. Consider $\boldsymbol{x} \in [0,1]^n \cap \mathbb{Q}^n$ such that $\boldsymbol{x} \neq \boldsymbol{v}$. The set $h_{\boldsymbol{v}}(\boldsymbol{x})$ can be expressed as an intersection of linear inequalities with rational coefficients, and is thus weakly recognizable in all bases, as a consequence of Theorems 1 and 2. From Lemma 1, it follows that the set $S' = S \cap h_{\boldsymbol{v}}(\boldsymbol{x})$ is recognizable both in bases r and s, by the same type of automaton as S. Besides, S' is both r^l- and $s^{l'}$-product-stable with respect to the pivot $\boldsymbol{v}$.

Let $C \in \mathbb{Q}^{n \times n}$ be a non-singular matrix such that $C(\boldsymbol{x} - \boldsymbol{v}) = (1, 0, \ldots, 0)$. Note that the transformation $\boldsymbol{y} \mapsto C\boldsymbol{y}$ maps $h_{\boldsymbol{v}}(\boldsymbol{x})$ onto a line segment that is parallel to the first axis. From Lemmas 1 and 2, we have that the set $S'' = C(S' - \boldsymbol{v}) \cap [0,1]^n$ inherits the recognizability properties of S. Moreover, S'' is r^l- and $s^{l'}$-product-stable with respect to the pivot $\mathbf{0}$.

Note that the set S'' can be decomposed into $S'' = S''' \times \{0\}^{n-1}$, with $S''' \subseteq [0,1]$. Applying Lemma 3, the set S''' has the same recognizability properties as S hence, by the inductive hypotheses, it is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. In other words, S''' is equal to a finite union of intervals with rational boundaries.

In addition, we know that S''' is r^l- and $s^{l'}$-product-stable with respect to the pivot 0. Since r and s are multiplicatively independent, r^l and $s^{l'}$ are multiplicatively independent as well. By Kronecker's approximation lemma, the set $\{r^{li} s^{l'j} \in]0,1] \mid i, j \in \mathbb{Z}\}$ is dense in $]0,1]$, as shown in [19,10]. It follows that if $1 \in S'''$ then $S''' \setminus \{0\} =]0,1]$ and if $1 \notin S'''$ then $S''' \setminus \{0\} = \emptyset$. As a consequence, $h_{\boldsymbol{v}}(\boldsymbol{x}) \cap S$ is either empty, or equal to $h_{\boldsymbol{v}}(\boldsymbol{x})$. □

Intuitively, Property 1 hints at the fact that the set S has a conical structure. We formalize this property by the following definition.

Definition 2. *A set $T \subseteq [0,1]^n$ is a* conical set *of vertex $\boldsymbol{v} \in [0,1]^n$ iff $\forall \boldsymbol{x} \in [0,1]^n, f \in]0,1] : \boldsymbol{x} \in T \Leftrightarrow f(\boldsymbol{x} - \boldsymbol{v}) + \boldsymbol{v} \in T$.*

In other words, a conical set is entirely determined by its vertex, and its intersection with the faces of the hypercube $[0,1]^n$. It follows that, in order to establish that S is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$, it suffices to show that this intersection is definable in the same theory, and that S is a conical set. We have the following results.

Property 2. For each $i \in \{1, 2, \ldots, n\}$ and $\lambda \in \{0, 1\}$, let $F_{\lambda,i} = \{\boldsymbol{x} \in [0,1]^n \mid x_i = \lambda\}$. The set $S \cap F_{\lambda,i}$ is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$.

[3] Note that this constraint implies that r and s are multiplicatively independent.

Proof sketch. Let $i \in \{1, 2, \ldots, n\}$ and $\lambda \in \{0, 1\}$. We first build the permutation matrix $C \in \{0, 1\}^{n \times n}$ such that $C\boldsymbol{x} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n, x_i)$ for any $\boldsymbol{x} = (x_1, \ldots, x_n) \in \mathbb{R}^n$. The set $F_{\lambda,i}$ is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$, hence it is weakly r-recognizable. From Lemmas 1 and 2, the set $S' = C(S \cap F_{\lambda,i})$ inherits the recognizability properties of S. Moreover, we have $S' = S'' \times \{\lambda\}$, where $S'' \subseteq [0, 1]^{n-1}$ has the same recognizability properties as well, as a consequence of Lemma 3. The result then follows from the inductive hypotheses. □

Property 3. The set S is conical with respect to the vertex $\boldsymbol{v}$.

Proof sketch. Let S' be the intersection of S with the faces of the hypercube $[0, 1]^n$. We have established that S' is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. The set S' can thus be expressed as a finite Boolean combination of linear constraints with rational coefficients. As a consequence, there exists a set $S'' \subseteq [0, 1]^n$ that is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$, conical with respect to the vertex $\boldsymbol{v}$, and that coincides with S' over the faces of $[0, 1]^n$ and over the vertex $\boldsymbol{v}$. Applying Property 1, we obtain $S'' \cap \mathbb{Q}^n = S \cap \mathbb{Q}^n$. From Theorem 4, we then have $S = S''$. □

5 Internal Structure of RVA

In Section 4, we have proved Theorems 5 and 6, which broadly state that if a set $S \subseteq [0, 1]^n$ is recognizable or weakly recognizable in two bases r and s that are sufficiently different, then this set is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. As a corollary, such sets are then weakly recognizable in every base $r \in \mathbb{N}_{>1}$ [11]. This result is significant, since it establishes that the class of weak deterministic automata is sufficient for representing all the sets that are recognizable by RVA regardless of the numeration base. As mentioned earlier, the advantage of using weak deterministic automata in actual applications comes from the fact that these automata are basically as easy to handle algorithmically as finite-word ones [20].

We now use Theorems 5 and 6, together with other results obtained in Section 4, in order to get some insight into the internal structure of RVA recognizing the subsets of $[0, 1]^n$ definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. As explained in Section 3, this is equivalent to studying the structure of RVA recognizing sets definable in $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, staying within the part of automata that reads the fractional part of vectors (i.e., in the sub-automata whose initial states are the destinations of transitions labeled by "$\star$").

Let $n \in \mathbb{N}_{>0}$ be a dimension, $r \in \mathbb{N}_{>1}$ be a base, and $S \subseteq [0, 1]^n$ be a set definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. We consider a weak and deterministic RVA $\mathcal{A}$ recognizing S in base r. We assume w.l.o.g. that $\mathcal{A}$ is complete as well as minimal (in the sense of [14]).

As observed in Section 3.2, for each state q of $\mathcal{A}$, the set $S(q) \subseteq [0, 1]^n$ encoded by the language $L(q)$ accepted from q can be derived from S by a transformation that is definable in $\langle \mathbb{R}, +, \leq, 1 \rangle$. It follows that each $S(q)$ is itself definable in that theory.

In addition, it has been established in Section 4.2 that, for the states q that belong to non-trivial strongly connected components of $\mathcal{A}$ (i.e., such that there exists at least one cycle from q to itself), the set $S(q)$ is a conical set. It is however worth noticing that the vertex of this conical set may not be uniquely determined. For instance, every element of the conical set $\{(0, \lambda) \mid \lambda \in [0, 1]\}$ is one of its vertices. We have the following result.

Theorem 7. *Let $n \in \mathbb{N}_{>0}$, and $T \subseteq [0, 1]^n$ be a conical set. The vertices of T form a bounded affine space $\{\boldsymbol{v} + \nu_1 \boldsymbol{u_1} + \cdots + \nu_m \boldsymbol{u_m} \in [0, 1]^n \mid \nu_1, \ldots, \nu_m \in \mathbb{R}\}$, with $m \in \mathbb{N}$ and $\boldsymbol{v}, \boldsymbol{u_1}, \ldots, \boldsymbol{u_m} \in \mathbb{R}^n$.*

Proof sketch. It is sufficient to prove that, if $\boldsymbol{v_1}, \boldsymbol{v_2} \in [0, 1]^n$ are two distinct vertices of T, then each point on the line segment $L = \{\mu \boldsymbol{v_1} + (1 - \mu) \boldsymbol{v_2} \in [0, 1]^n \mid \mu \in \mathbb{R}\}$ linking $\boldsymbol{v_1}$ and $\boldsymbol{v_2}$ is also a vertex of T. This can be achieved by showing that T is invariant under any translation parallel to L that stays within $[0, 1]^n$, i.e., that $\boldsymbol{x} \in T \Leftrightarrow \boldsymbol{x} + \mu(\boldsymbol{v_1} - \boldsymbol{v_2}) \in T$ for all $\boldsymbol{x} \in [0, 1]^n$ and $\mu \in \mathbb{R}$ such that $\boldsymbol{x} + \mu(\boldsymbol{v_1} - \boldsymbol{v_2}) \in [0, 1]^n$.

Let $\boldsymbol{x}$ be an arbitrary vector in $[0, 1]^n$. Consider an arbitrary value $\mu \in \mathbb{R}_{\geq 0}$ such that $\boldsymbol{x} + \mu(\boldsymbol{v_1} - \boldsymbol{v_2}) \in [0, 1]^n$. (Note that restricting μ to be non negative does not weaken the property.)

We define $\boldsymbol{x'} = \boldsymbol{x} + \mu(\boldsymbol{v_1} - \boldsymbol{v_2})$ and $f = 1/(1 + \mu)$. Since T is a conical set w.r.t. the vertex $\boldsymbol{v_1}$, we have $\boldsymbol{x} \in T$ if and only if $f(\boldsymbol{x} - \boldsymbol{v_1}) + \boldsymbol{v_1} \in T$. Exploiting the conical structure of T w.r.t. the vertex $\boldsymbol{v_2}$, we then get $\boldsymbol{x'} \in T$ if and only if $f(\boldsymbol{x'} - \boldsymbol{v_2}) + \boldsymbol{v_2} \in T$. By replacing $\boldsymbol{x'}$ by $\boldsymbol{x} + \mu(\boldsymbol{v_1} - \boldsymbol{v_2})$ we deduce the equality

$$f(\boldsymbol{x'} - \boldsymbol{v_2}) + \boldsymbol{v_2} \ = \ f(\boldsymbol{x} - \boldsymbol{v_1}) + \boldsymbol{v_1}$$

which yields $\boldsymbol{x} \in T \Leftrightarrow \boldsymbol{x'} \in T$. □

We are now ready to describe the structure of $\mathcal{A}$: Its initial state is the root of a (possibly empty) acyclic structure, composed of states belonging to trivial strongly connected components, and leading to states q belonging to non-trivial components. For each such state q, the set $S(q) \subseteq [0, 1]^n$ is conical. Such a set is entirely characterized by the affine space containing its vertices, a Boolean value stating whether these vertices belong or not to $S(q)$, and the intersection of the set with the $2n$ faces of the hypercube $[0, 1]^n$. This intersection can be expressed in terms of at most $2n$ subsets of $[0, 1]^{n-1}$ (in the bottom case $n = 1$, this degenerates into the two extremities of the interval $[0, 1]$), which are known to be recognizable in $\langle \mathbb{R}, +, \leq, 1 \rangle$.

Those observations could lead to a more efficient data structure for representing canonically the subsets of $[0, 1]^n$ that are definable in additive arithmetic. For instance, RVA could be represented implicitly, by using BDDs for describing their initial acyclic structure [21] and linking this structure to representations of conical sets. These sets could be described by encoding separately the affine space containing their vertices and the faces of their enclosing hypercube. These faces could themselves be represented by the same type of structure applied to subsets of $[0, 1]^{n-1}$. The bottom layer of such a hierarchical representation would correspond to individual rational numbers, which could be encoded explicitly.

The detailed study of such a representation system, and its application to decision procedures for $\langle \mathbb{R}, +, \leq, 1 \rangle$ and $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$, will be the subject of future work.

6 Conclusions

In this paper, we have characterized the subsets of $\mathbb{R}^n$, with $n \in \mathbb{N}_{>0}$, that are recognizable by RVA, or weak deterministic RVA, in multiple bases. Precisely, we have established that the sets that are either weakly recognizable in two multiplicatively independent bases, or recognizable in two bases that do not share the same set of prime factors, are exactly those that are definable in the first-order theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. These results were already known for the particular case $n = 1$ [17,10]. They generalize to automata operating on real vectors Semenov's theorem, which states that the sets of integer vector that are recognizable in multiplicatively independent bases are those that are definable in Presburger arithmetic. The theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$ can indeed be seen as an extension of Presburger arithmetic to mixed integer and real variables [22]. It is worth mentioning that, in the case of (non weak) recognizability, the condition on the numeration bases cannot be replaced by multiplicative independence. Indeed, there exist subsets of $\mathbb{R}$ that are simultaneously recognizable in two multiplicatively independent bases, but without being definable in additive arithmetic [10].

An important corollary of our results is that every subset of $\mathbb{R}^n$ that is recognizable in every base $r \in \mathbb{N}_{>1}$ can be recognized by a weak deterministic automaton. This provides a theoretical justification to the use of these automata for representing sets of integer and real vectors, in addition to their practical advantages.

As an additional contribution, we have obtained interesting insight into the structure of weak deterministic automata recognizing sets definable in the theory $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. In future work, we will address the problem of exploiting this structure in order to develop more efficient symbolic representation systems for subsets of $\mathbb{R}^n$, as well as an improved decision procedure for $\langle \mathbb{R}, \mathbb{Z}, +, \leq \rangle$. Our aim is to be able to benefit from the advantages of automata-based symbolic representations, which mainly reside in their easy algorithmic manipulation and their canonicity, while managing to avoid some of their drawbacks, such as the unnecessarily large size of automata obtained from some classes of constraints [23]. This could be achievable by keeping a part of the transition relation of RVA implicit. Such a representation would also simplify the problem of extracting formulas from automata recognizing arithmetic sets [24,25].

References

1. Presburger, M.: Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In: Comptes Rendus du Premier Congrès des Mathématiciens des Pays Slaves, Warsaw (1929) 92–101

2. Büchi, J.R.: On a decision method in restricted second order arithmetic. In: Proc. International Congress on Logic, Methodoloy and Philosophy of Science, Stanford, Stanford University Press (1962) 1–12
3. Bruyère, V., Hansel, G., Michaux, C., Villemaire, R.: Logic and p-recognizable sets of integers. Bulletin of the Belgian Mathematical Society **1**(2) (1994) 191–238
4. Boigelot, B.: Symbolic methods for exploring infinite state spaces. PhD thesis, Université de Liège (1998)
5. Cobham, A.: On the base-dependence of sets of numbers recognizable by finite automata. Mathematical Systems Theory **3** (1969) 186–192
6. Semenov, A.: Presburgerness of predicates regular in two number systems. Siberian Mathematical Journal **18** (1977) 289–299
7. Boigelot, B., Bronne, L., Rassart, S.: An improved reachability analysis method for strongly linear hybrid systems. In: Proc. 9th CAV. Volume 1254 of Lecture Notes in Computer Science., Haifa, Springer (1997) 167–177
8. The Liège Automata-based Symbolic Handler (LASH). Toolset available at: `http://www.montefiore.ulg.ac.be/~boigelot/research/lash/`
9. Becker, B., Dax, C., Eisinger, J., Klaedtke, F.: LIRA: Handling constraints of linear arithmetics over the integers and the reals. In: Proc. 19th CAV. Volume 4590 of Lecture Notes in Computer Science., Springer (2007) 307–310
10. Boigelot, B., Brusten, J., Bruyère, V.: On the sets of real numbers recognized by finite automata in multiple bases. In: Proc. 35th. ICALP. Volume 5126 of Lecture Notes in Computer Science., Springer (2008) 112–123
11. Boigelot, B., Jodogne, S., Wolper, P.: An effective decision procedure for linear arithmetic over the integers and reals. ACM Transactions on Computational Logic **6**(3) (2005) 614–633
12. van Leeuwen, J., ed.: Handbook of Theoretical Computer Science, Volume B: Formal Models and Semantics. Elsevier and MIT Press (1990)
13. Vardi, M.: The Büchi complementation saga. In: Proc. 24th. STACS. Volume 4393 of Lecture Notes in Computer Science., Aachen, Springer (2007) 12–22
14. Löding, C.: Efficient minimization of deterministic weak ω-automata. Information Processing Letters **79**(3) (2001) 105–109
15. Boigelot, B., Rassart, S., Wolper, P.: On the expressiveness of real and integer arithmetic automata. In: Proc. 25th ICALP. Volume 1443 of Lecture Notes in Computer Science., Aalborg, Springer (1998) 152–163
16. Perrin, D., Pin, J.: Infinite words. Volume 141 of Pure and Applied Mathematics. Elsevier (2004)
17. Boigelot, B., Brusten, J.: A generalization of Cobham's theorem to automata over real numbers. Theoretical Computer Science **In Press** (2009)
18. Ferrante, J., Rackoff, C.: The Computational Complexity of Logical Theories. Volume 718 of Lecture Notes in Mathematics. Springer (1979)
19. Hardy, G.H., Wright, E.M.: An introduction to the theory of numbers. 5th edn. Oxford University Press (1985)
20. Wilke, T.: Locally threshold testable languages of infinite words. In: Proc. 10th STACS. Volume 665 of Lecture Notes in Computer Science., Würzburg, Springer (1993) 607–616
21. Bryant, R.: Symbolic Boolean manipulation with ordered binary decision diagrams. ACM Computing Surveys **24**(3) (1992) 293–318
22. Weispfenning, V.: Mixed real-integer linear quantifier elimination. In: Proc. ACM SIGSAM ISSAC, Vancouver, ACM Press (1999) 129–136

23. Eisinger, J., Klaedtke, F.: Don't care words with an application to the automata-based approach for real addition. In: Proc. 18th CAV. Volume 4144 of Lecture Notes in Computer Science., Seattle, Springer (2006) 67–80
24. Latour, L.: Presburger arithmetic: from automata to formulas. PhD thesis, Université de Liège (2005)
25. Leroux, J.: A polynomial time Presburger criterion and synthesis for number decision diagrams. In: Proc. 20th LICS, Chicago, IEEE Computer Society (2005) 147–156