Quantum finite automata and linear context-free languages: a decidable problem

Alberto Bertoni[1], Christian Choffrut[2], and Flavio D'Alessandro[3]

1 **Dipartimento di Scienze dell'Informazione, Università degli Studi di Milano**
Via Comelico 39, 20135 Milano, Italy
bertoni@dsi.unimi.it

2 **Laboratoire LIAFA, Université de Paris 7**
2, pl. Jussieu, 75251 Paris Cedex 05
cc@liafa.jussieu.fr

3 **Dipartimento di Matematica, "La Sapienza" Università di Roma**
Piazzale Aldo Moro 2, 00185 Roma, Italy
dalessan@mat.uniroma1.it

Abstract

We consider the so-called measure once finite quantum automata model introduced by Moore and Crutchfield in 2000. We show that given a language recognized by such a device and a linear context-free language, it is recursively decidable whether or not they have a nonempty intersection. This extends a result of Blondel et al. which can be interpreted as solving the problem with the free monoid in place of the family of linear context-free languages.

1998 ACM Subject Classification 81P68, 68Q45, 20G20

Keywords and phrases Quantum automata, Context-free languages, Algebraic groups, Decidability

Digital Object Identifier 10.4230/LIPIcs.xxx.yyy.p

1 Introduction

Quantum finite automata or simply quantum automata were introduced at the beginning of the previous decade in [6] as a new model of language recognizer. Numerous publications have ever since compared their decision properties to those of the older model of probabilistic finite automata. Some undecidable problems for probabilistic finite automata turn out to be decidable for quantum finite automata. The result in [3] which triggered our investigation can be viewed as asserting that the intersection emptiness problem of a language recognized by a finite quantum automaton with the free monoid is recursively decidable. The present result concerns the same problem where instead of the free monoid, more generally a language belonging to some classical families of languages such as the context-free languages and the bounded semilinear languages is considered.

An ingredient of the proof in [3] consists of expressing the emptiness problem in the first order theory of the reals and then to apply Tarski-Seidenberg quantifier elimination. This is possible because an algebraic subset, i.e., a closed subset in the Zariski topology $\mathcal{A} \subseteq \mathbb{R}^n$, is naturally associated to this intersection and even more miraculously because this subset can be effectively computed (cf. also [4]).

Here we show that the (actually semi-)algebraicity of $\mathcal{A}$ still holds when considering not only the free monoid but more generally arbitrary context-free languages and bounded semilinear languages. Unfortunately, its effective construction is only guaranteed under

stricter conditions such as the fact that the language is context-free and linear or is bounded semilinear. In particular, in the case of context-free languages, we are not able to settle the nonlinear case yet.

We now give a more formal presentation of our work. The free monoid generated by the finite alphabet Σ is denoted by Σ^*. The elements of Σ^* are *words*. We consider all finite dimensional vector spaces as provided with the Euclidian norm. A *quantum automaton* is a quadruple $\mathcal{Q} = (s, \varphi, P, \lambda)$ where $s \in \mathbb{R}^n$ is a row-vector of unit norm, P is a projection of $\mathbb{R}^n$, φ is a representation of the free monoid Σ^* into the group of *orthogonal* $n \times n$-matrices in $\mathbb{R}^{n\times n}$ and the *threshold* λ has value in $\mathbb{R}$. We recall that a real matrix M is orthogonal if its inverse equals its transpose: $M^{-1} = M^T$. We denote by O_n the group of $n \times n$-orthogonal matrices. We are mainly interested in effective properties which requires the quantum automaton to be effectively given. We say that the quantum automaton is *rational* if all the coefficients of the components of the automaton are rational numbers, i.e., φ maps Σ^* into $\mathbb{Q}^{n\times n}$ and $\lambda \in \mathbb{Q}$. This hypothesis is not a restriction since all we use for the proofs is the fact that the arithmetic operations and the comparison are effective in the field of rational numbers. This is the "measure once" model introduced by Moore and Crutchfield in 2000 [6]. The language recognized by $\mathcal{Q}$ is

$$||\mathcal{Q}_>|| = \{w \in \Sigma^* \mid ||s\varphi(w)P|| > \lambda\} \tag{1}$$

Blondel et al. in [3] proved that the emptiness problem of $||\mathcal{Q}_>||$ is decidable. This can be interpreted as saying that the emptiness problem of the intersection of a language accepted by a quantum automaton and the specific language Σ^* is decidable. In other word, it falls into the category of issues asking for the decision status of the intersection of two languages. It is known that such a problem is already undecidable at a very low level of the complexity hierarchy of recursive languages, namely for linear context-free languages to which Post Correspondence Problem can be easily reduced.

A few words on the technique used in the above paper. Observe that, with the natural meaning of the notation $||\mathcal{Q}_\leq||$, the emptiness problem for languages defined by (1) is equivalent to the inclusion

$$\Sigma^* \subseteq ||\mathcal{Q}_\leq|| \tag{2}$$

Since the function $M \to ||sMP||$ is continuous, it is sufficient to prove that for all matrices M in the tolopological closure of $\varphi(\Sigma^*)$ the condition $||sMP|| \leq \lambda$ holds. The nonemptiness is clearly semidecidable. In order to prove that the emptiness is semidecidable the authors resort to two ingredients. They observe that the topological closure of the monoid of matrices $\varphi(\Sigma^*)$ is algebraic, i.e., when considering the $n \times n$-entries of a matrix M in the topological closure of $\varphi(\Sigma^*)$ as as many unknowns in the field of reals, they are precisely the zeros of a polynomial in $\mathbb{R}[x_{1,1}, \ldots, x_{n,n}]$. This allows them to express the property (2) in first-order logic of the field of reals. The second ingredient consists of applying Tarski-Seidenberg quantifier elimination and Hilbert basis results, which yields decidability.

We generalize the problem by considering families of languages $\mathcal{L}$ instead of the fixed language Σ^*. The question we tackle is thus as follows

L-Q INTERSECTION

INPUT: a language L in a family of languages $\mathcal{L}$ and a finite quantum automaton $\mathcal{Q}$.

QUESTION: does $L \cap ||\mathcal{Q}_>|| = \emptyset$ hold?

Our main result shows that whenever $\mathcal{L}$ is the collection of linear context-free languages or is the collection of bounded semilinear languages, and whenever the automaton is rational, the problem is decidable. It can be achieved, not only because the orthogonal matrices associated with L are semialgebraic (a more general property than algebraic, which is defined by more general first-order formulas), but also because these formulas can be computed "in the limit".

We can prove the semialgebraicity of more general families of languages: arbitrary subsemigroups which is a trivial case and context-free languages which is less immediate.

In the last section, we show that our main results are not trivial since we can exhibit an example of a language which is the complement of a linear context-free language and whose set of matrices is not semialgebraic.

2 Preliminaries

Throughout this paper the notation (s, φ, P, λ) stands for a quantum automaton $\mathcal{Q}$ where, as mentioned in the Introduction, $s \in \mathbb{R}^n$ is a vector of unit norm, P is a projection of $\mathbb{R}^n$, φ is a representation of the free monoid Σ^* into the group O_n of orthogonal $n \times n$-matrices in $\mathbb{R}^{n\times n}$. The behaviour of $\mathcal{Q}$ heavily depends on the topological properties of the semigroup of matrices $\varphi(\Sigma^*)$. This is why, before returning to quantum automata, we first focus our attention on these matrices for their own sake.

2.1 Topology

The following result is needed in the proof of the main theorem. Though valid under weaker conditions, it will be considered in the particular case of orthogonal matrices. Given a subset E of a vector space, we denote by $\mathbf{C}\mathrm{l}(E)$ the topological closure for the topology induced by the Euclidian norm. Given a k-tuple of matrices $(M_1, \ldots, M_k)$, denote by f the k-ary product $f(M_1, \ldots, M_k) = M_1 \cdots M_k$ and extend the notation to subsets ρ of k-tuples of matrices by posing $f(\rho) = \{f(M_1, \ldots, M_k) \mid M_1 \cdots M_k \in \rho\}$. The following result will be applied in several instances of this paper. It says that because we are dealing with compact subsets, the two operators of matrix multiplication and the topological closure commute (For the proof of Theorem 1, see the Appendix).

▶ **Theorem 1.** *Let $\mathcal{C}$ be a compact subset of matrices and let $\rho \subseteq \mathcal{C}^k$ be a k-ary relation. Then we have*

$$\mathbf{C}l(f(\rho)) = f(\mathbf{C}l(\rho))$$

Consequently, if ρ is a binary relation which is a direct product $\rho_1 \times \rho_2$, we have $\mathbf{C}l(\rho_1\rho_2) = f(\mathbf{C}l(\rho_1 \times \rho_2))$. It is an elementary result of functional analysis that $\mathbf{C}l(\rho_1 \times \rho_2) = \mathbf{C}l(\rho_1) \times \mathbf{C}l(\rho_2)$ holds. Because of $\mathbf{C}l(\rho_1\rho_2) = f(\mathbf{C}l(\rho_1 \times \rho_2)) = f(\mathbf{C}l(\rho_1) \times \mathbf{C}l(\rho_2)) = \mathbf{C}l(\rho_1)\ \mathbf{C}l(\rho_2)$ we have

▶ **Corollary 2.** *The topological closure of the product of two sets of matrices included in a compact subspace is equal to the product of the topological closures of the two sets.*

2.2 Algebraic and semialgebraic sets

Let us give first the definition of algebraic set over the field of real numbers (cf. [2, 7]).

▶ **Definition 3.** A subset $\mathcal{A} \subseteq \mathbb{R}^n$ is *algebraic (over the field of real numbers)*, if it satisfies one of the following equivalent conditions:

(i) $\mathcal{A}$ is the zero set of a polynomial $p \in \mathbb{R}[x_1, \ldots, x_n]$, i.e.,

$$v \in \mathcal{A} \iff p(v) = 0. \tag{3}$$

(ii) $\mathcal{A}$ is the zero set of an arbitrary set of polynomials $\mathcal{P}$ with coefficients in $\mathbb{R}[x_1, \ldots, x_n]$, i.e., for every vector $v \in \mathbb{R}^n$,

$$v \in \mathcal{A} \iff \forall\, p \in \mathcal{P} :\ p(v) = 0. \tag{4}$$

The equivalence of the two statements is a consequence of Hilbert finite basis Theorem. Indeed, it claims that given a collection $\mathcal{P}$ there exists a finite subcollection $p_1, \ldots, p_r$ generating the same ideal which implies in particular that for all $p \in \mathcal{P}$ there exist $q_1, \ldots, q_r$ with

$$p = q_1 p_1 + \cdots + q_r p_r$$

Then $p_j(v) = 0$ for $j = 1, \ldots, r$ implies $p(v) = 0$. Now this finite set of equations can be reduced to the unique equation

$$\sum_{i=1}^{n} p_j(x)^2 = 0$$

As a trivial example, a singleton $\{v\}$ is algebraic since it is the unique solution of the equation

$$\sum_{i=1}^{n} (x_i - v_i)^2 = 0$$

where v_i, with $1 \leq i \leq n$, is the i-th component of the vector v.

It is routine to check that the family of algebraic sets is closed under finite unions and intersections. However, it is not closed under complement. The following more general class of subsets enjoys extra closure properties and is therefore more robust. The equivalence of the two definitions below is guaranteed by Tarski-Seidenberg quantifier elimination result.

▶ **Definition 4.** A subset $\mathcal{A} \subseteq \mathbb{R}^n$ is *semialgebraic (over the field of real numbers)* if it satisfies one of the two equivalent conditions

(i) $\mathcal{A}$ is the set of vectors satisfying a finite Boolean combination of predicates of the form $p(x_1, \ldots, x_n) > 0$ where $p \in \mathbb{R}[x_1, \ldots, x_n]$.

(ii) $\mathcal{A}$ is first-order definable on the structure whose domain are the reals and whose predicates are of the form $p(x_1, \ldots, x_n) > 0$ and $p(x_1, \ldots, x_n) = 0$ with $p \in \mathbb{R}[x_1, \ldots, x_n]$.

We now specify these definitions to square matrices.

▶ **Definition 5.** A set $\mathcal{A} \subseteq \mathbb{R}^{n \times n}$ of matrices is *algebraic*, resp. *semialgebraic*, if considered as a set of vectors, it is algebraic, resp. semialgebraic.

We now combine the notions of zero sets and of topology. In the following two results we rephrase Theorem 3.1 of [3] by emphasizing the main features that serve our purpose. Given a subset E of a group, we denote by $\langle E \rangle$ and by E^* the subgroup and the submonoid it generates, respectively.

▶ **Theorem 6.** *Let $S \subseteq \mathbb{R}^{n \times n}$ be a set of orthogonal matrices and let E be any subset satisfying $\langle S \rangle = \langle E \rangle$. Then we have $\mathbf{C}l(S^*) = \mathbf{C}l(\langle E \rangle)$.*

Proof. It is known that every compact subsemigroup of a compact group is a subgroup G. Now $S^* \subseteq \langle E \rangle$ implies $G = \mathbf{C}l(S^*) \subseteq \mathbf{C}l(\langle E \rangle)$ and $S \subseteq G$ implies $\mathbf{C}l(\langle E \rangle) \subseteq G$ and thus $\mathbf{C}l(S^*) = \mathbf{C}l(\langle E \rangle)$. ◀

The main consequence of the next theorem is that the topological closure of a monoid of orthogonal matrices is algebraic (for the proof of this theorem, see the Appendix).

▶ **Theorem 7.** *Let E be a set of orthogonal matrices. Then $\mathbf{C}l(\langle E \rangle)$ is a subgroup of orthogonal matrices and it is the zero set of all polynomials $p[x_{1,1}, \ldots, x_{n,n}]$ satisfying the conditions*

$$p(I) = 0 \quad \textit{and } p(eX) = p(X) \quad \textit{for all } e \in E$$

Furthermore, if the matrices in E have rational coefficients, the above condition may be restricted to polynomials with coefficients in $\mathbb{Q}$.

Combining the previous two theorems, we get the general result

▶ **Corollary 8.** *Let $L \subseteq \Sigma^*$. Then $\mathbf{C}l(\varphi(L)^*)$ is algebraic.*

2.3 Effectiveness issues

We now return to the L-Q INTERSECTION problem as defined in the Introduction. We want to prove the implication

$$\forall X : X \in \varphi(L) \Rightarrow ||sXP|| \leq \lambda$$

We observed that due to the fact that the function $X \to ||sXP||$ is continuous the implication is equivalent to the implication

$$\forall X : X \in \mathbf{C}l(\varphi(L)) \Rightarrow ||sXP|| \leq \lambda$$

It just happens that under certain hypotheses, $\mathbf{C}l(\varphi(L))$ is semialgebraic, i.e., it is defined by a first-order formula which turns the above statement into a first order formula. In the simplest examples, the closure is defined by an infinite conjunctions of equations which by Hilbert finite basis result reduces to a unique equation. Thus Theorerm 7 guarantees the existence of the formula but does not give an upper bound on the finite number of equations which must be tested. Therefore the following definition is instrumental for the rest of the paper.

▶ **Definition 9.** A subset $\mathcal{A}$ of matrices is *effectively eventually definable* if there exists a constructible sequence of first-order formulas ϕ_i satisfying the conditions

1) for all $i \geq 0$, for all $X \in \mathbb{R}^{n \times n}$ we have $\phi_{i+1}(X) \Rightarrow \phi_i(X)$

2) for all $i \geq 0$, for all $X \in \mathcal{A}$ we have $\phi_i(X)$

3) there exists an integer n such that

$$\phi_n(X) \Rightarrow X \in \mathcal{A}$$

The following is a first application of the notion and illustrates the discussion before the definition.

▶ Proposition 1. Let $\mathcal{Q}$ be a rational quantum automaton. Let $L \subseteq \Sigma^*$ be such that the set $\mathbf{C}l(\varphi(L))$ is effectively eventually definable. It is recursively decidable whether or not $L \cap ||\mathcal{Q}_>|| = \emptyset$ holds.

Proof. Equivalently we prove the inclusion $L \subseteq ||\mathcal{Q}_{\leq}||$. In order to prove that the inclusion is effective, we proceed as in [3]. We run in parallel two semialgorithms. The first one verifies the noninclusion by enumerating the words $w \in L$ and testing if $||s\varphi(w)P|| > \lambda$ holds. The second semialgorithm considers a sequence of formulas $\phi_i(X)$, $i = 0, \ldots$, which effectively eventually defines $\mathbf{C}l(\varphi(L))$ and verifies whether the sentence

$$\Psi_i \equiv \forall X : \phi_i(X) \Rightarrow sXP \leq \lambda$$

holds which can be achieved by Tarski Seidenberg elimination result. If the inclusion $L \subseteq ||\mathcal{Q}_{\leq}||$ holds then the first semialgorithm cannot answer "yes" and the second semialgorithm will eventually answer "yes". If the inclusion does not hold then the second semialgorithm cannot answer "yes" for any Ψ_i since the second condition of the Definition 9 implies $X \in \mathbf{C}l(\varphi(L)) \Rightarrow \phi_i(X)$ and thus

$$\forall X : X \in \mathbf{C}l(\varphi(L)) \Rightarrow ||sXP|| \leq \lambda$$

a contradiction. ◀

We state a sufficient condition for a subset of matrices to be effectively eventually definable.

Let $S \subseteq \mathbb{R}^{n\times n}$ be a set of orthogonal matrices and let E be any subset satisfying $\langle S\rangle = \langle E\rangle$.

▶ Proposition 2. Let $L \subseteq \Sigma^*$ and let $E \subseteq \mathbb{Q}^{n\times n}$ be a finite subset of orthogonal matrices satisfying $\langle\varphi(L)\rangle = \langle E\rangle$. Then $\mathbf{C}l(\varphi(L)^*)$ is effectively eventually definable.

Proof. Indeed, set $\mathcal{A} = \mathbf{C}l(\varphi(L)^*) = \mathbf{C}l(\langle E\rangle)$ where the last equality is guaranteed by Theorem 6. Then $\mathcal{A}$ is the zero set of all polynomials $p(X)$ where p satisfies the condition

$$p(I) = 0 \text{ and } p(gX) = p(X) \text{ for all } g \in \mathcal{A}$$

Since it clearly suffices to verify the invariance of p under the action of the finite set of generators, we proceed as follows. We enumerate all polynomials $p \in \mathbb{Q}[x_{1,1}, \ldots, x_{n,n}]$ say $p_0, p_1, \ldots$. For each such polynomial p the invariance relative to the action of each generator can be tested. Thus the formula

$$\phi_i(X) \equiv \bigwedge_j^i p_j(X) = 0$$

effectively eventually defines $\mathcal{A}$: the first two conditions can be readily verified and the last one is a consequence of Hilbert finite basis theorem on ideals of polynomials. ◀

2.4 Closure properties

In this paragraph we investigate some closure properties of the three different classes of matrices, algebraic, semialgebraic and effectively eventually definable, under the main usual operations as well as new operations.

We define the *sandwich* operation denoted by $\diamond$ whose first operand is a set of pairs of matrices $\mathcal{A} \subseteq \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ and the second operand a set of matrices $\mathcal{B} \subseteq \mathbb{R}^{n\times n}$ by setting

$$\mathcal{A} \diamond \mathcal{B} = \{XYZ \mid (X, Z) \in \mathcal{A} \text{ and } Y \in \mathcal{B}\}$$

The next operation will be used. Given a bijection

$$\pi : \{(i, j) \mid i, j \in \{1, \ldots, n\}\} \to \{(i, j) \mid i, j \in \{1, \ldots, n\}\} \tag{5}$$

and a matrix $M \in \mathbb{R}^{n\times n}$ denote by $\pi(M)$ the matrix $\pi(M)_{i,j} = M_{\pi(i,j)}$. Extend this operation to subsets of matrices $\mathcal{A}$. Write $\pi(\mathcal{A})$ to denote the set of matrices $\pi(M)$ for all $M \in \mathcal{A}$.

The last operation is the *sum* of square matrices $M_1, \ldots, M_k$ whose result is the square block matrix

$$M_1 \oplus \cdots \oplus M_k = \begin{pmatrix} M_1 & 0 & \cdots & 0 \\ 0 & M_2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & M_k \end{pmatrix} \tag{6}$$

These notations extend to subsets of matrices in the natural way. Here we assume that all k matrices have the same dimension $n \times n$. Observe that if the matrices are orthogonal, so is their sum. Such matrices form a subgroup of orthogonal matrices of dimension $kn \times kn$.

Logic provides an elegant way to formulate properties in the present context. Some conventions are used throughout this work. E.g., we write $\exists^n X$ when we mean that X is a vector of n bound variables. Furthermore, a vector of $n \times n$ variables can be interpreted as an $n \times n$ matrix of variables. As a consequence of Tarski-Seidenberg result, consider two semialgebraic subsets of matrices, say $\mathcal{A}_1$ and $\mathcal{A}_2$, defined by two first-order formulas $\phi_1(X_1)$ and $\phi_2(X_2)$ where X_1 and X_2 are two collections of n^2 free variables viewed as two $n \times n$ matrices of variables. Then the product

$$\mathcal{A}_1\mathcal{A}_2 = \{M_1M_2 \mid M_1 \in \mathcal{A}_1, M_2 \in \mathcal{A}_2\}$$

is defined by the following formula where X is a collection of n^2 free variables viewed as an $n \times n$ matrix

$$\exists^{n\times n} X_1 \exists^{n\times n} X_2 : X = X_1X_2 \wedge \phi_1(X_1) \wedge \phi_2(X_2)$$

where $X = X_1X_2$ is an abbreviation for the predicate defining X as the matrix product of X_1 and X_2. This proves that the product of two semialgebraic sets of matrices is semialgebraic. Similarly we have the following closure properties whose verification is routine.

▶ Proposition 3. Let $\mathcal{A}_1, \mathcal{A}_2 \subseteq \mathbb{R}^{n\times n}$ be two sets of matrices and let π be a one-to-one mapping as in (5).

1) If $\mathcal{A}_1$ and $\mathcal{A}_2$ are algebraic so are $\mathcal{A}_1\mathcal{A}_2$ and $\pi(\mathcal{A}_1)$.

2) If $\mathcal{A}_1$ and $\mathcal{A}_2$ are semialgebraic, resp. effectively eventually definable, so are $\mathcal{A}_1 \cup \mathcal{A}_2$, $\mathcal{A}_1\mathcal{A}_2$ and $\pi(\mathcal{A}_1)$.

▶ Proposition 4. Let $\mathcal{A}_1 \subseteq \mathbb{R}^{n\times n} \times \mathbb{R}^{n\times n}$ and $\mathcal{A}_2 \subseteq \mathbb{R}^{n\times n}$ be semialgebraic, resp. effectively eventually definable. Then $\mathcal{A}_1 \diamond \mathcal{A}_2$ is semialgebraic, resp. effectively eventually definable.

▶ Proposition 5. Let $\mathcal{A}$ be a semialgebraic, resp. effectively eventually definable, set of $kn \times kn$ matrices of the form (6). The set

$$\{X_1 \cdots X_k \mid X_1 \oplus \cdots \oplus X_k \in \mathcal{A}\}$$

is semialgebraic, resp. effectively eventually definable.

▶ Proposition 6. If $\mathcal{A}_1, \ldots, \mathcal{A}_k \subseteq \mathbb{R}^{n\times n}$ are semialgebraic, resp. effectively eventually definable sets of matrices then so is the set $\mathcal{A}_1 \oplus \cdots \oplus \mathcal{A}_k$.

3 Context-free languages

For the sake of self-containment and in order to fix notation, we recall the basic properties and notions concerning the family of context-free languages which can be found in all introductory textbooks on theoretical computer science (see, for instance, [5]).

A *context-free grammar* G is a quadruple $\langle V, \Sigma, P, S\rangle$ where Σ is the alphabet of *terminal symbols*, V is the set of *variables*, P is the set of *rules*, and S is the *axiom* of the grammar. A word over the alphabet Σ is called *terminal*. As usual, the variables are denoted by uppercase letters A, B, ... A typical rule of the grammar is written as $A \to \alpha$. The *derivation* relation of G is denoted by $\stackrel{*}{\Rightarrow}$.

A grammar is *linear* if every right hand side α contains at most one occurrence of variables, i.e., if it belongs to $\Sigma^* \cup \Sigma^* V \Sigma^*$.

The idea of the following notation is to consider the set of all pairs of left and right contexts in the terminal alphabet of a self-embedding variable. In the next definition, the initial "C" is meant to suggest the term "context" as justified by the following.

▶ **Definition 10.** With each variable $A \in V$ associate its *terminal contexts* defined as

$$C_A = \{(\alpha, \beta) \in \Sigma^* \times \Sigma^* : A \stackrel{*}{\Rightarrow} \alpha A \beta\}.$$

As the proof of the main theorem proceeds by induction on the number of variables, we need to show how to recombine a grammar from simpler ones obtained by choosing an arbitrary non-axiom symbol as the new axiom and by canceling all the rules involving S. This is the reason for introducing the next notation

▶ **Definition 11.** Let $G = \langle V, \Sigma, P, S\rangle$ be a context-free grammar. Set $V' = V \setminus \{S\}$.

For every $A \in V'$, define the context-free grammar $G_A = \langle V', \Sigma, P_A, A\rangle$ where the set P_A consists of all the rules of G of the form

$$B \to \gamma, \quad B \in V', \quad \gamma \in (V' \cup \Sigma)^*$$

and denote by L_A the language of all terminal words generated by the grammar G_A.

The next result introduces the language of terminal words obtained in a derivation where S occurs at the start only.

▶ **Definition 12.** Let $L'(G)$ denote the set of all the words of Σ^* which admit a derivation

$$S \Rightarrow \gamma_1 \Rightarrow \cdots \Rightarrow \gamma_\ell \Rightarrow w \tag{7}$$

where, for every $i = 1, \ldots, \ell$, $\gamma_i \in (V' \cup \Sigma)^*$.

If no ambiguity arises, in the sequel, the language $L'(G)$ is simply denoted L'.

The language L' can be easily expressed in terms of the languages L_A for all $A \in V'$. Indeed, consider the set of all rules of the grammar G of the form

$$S \to \beta, \quad \beta \in (V' \cup \Sigma)^* \tag{8}$$

Factorize every such β as

$$\beta = w_1 A_1 w_2 A_2 \cdots w_\ell A_\ell w_{j_{\ell+1}} \tag{9}$$

where $w_1, \ldots, w_{\ell+1} \in \Sigma^*$ and $A_1, A_2, \ldots A_\ell \in V'$. The following is a standard exercise.

▶ **Lemma 13.** *With the notations of (9), the language L' is the (finite) union of the languages*

$$w_1 L_{A_1} w_2 L_{A_2} \cdots w_\ell L_{A_\ell} w_{j_{\ell+1}}$$

when β ranges over all rules (8).

▶ Proposition 7. With the previous notations L is a finite union of languages of the form $C_S \diamond L''$ where

$$L'' = w_1 L_{A_1} w_2 L_{A_2} \cdots w_\ell L_{A_\ell} w_{\ell+1}$$

Proof. In order to prove the inclusion of the right- into left- hand side, it suffices to consider $w = \alpha u \beta$, with $u \in L'$ and $(\alpha, \beta) \in C_S$. One has $S \stackrel{*}{\Rightarrow} u$ and $S \stackrel{*}{\Rightarrow} \alpha S \beta$ and thus $S \stackrel{*}{\Rightarrow} \alpha S \beta \stackrel{*}{\Rightarrow} \alpha u \beta$.

Let us prove the opposite inclusion. A word $w \in L$ admits a derivation $S \stackrel{*}{\Rightarrow} w$. If the symbol S does not occur in the derivation except at the start of the derivation, then $w \in L'$. Otherwise factor this derivation into $S \stackrel{*}{\Rightarrow} \alpha S \beta \stackrel{*}{\Rightarrow} w$ such that S does not occur in the second part of the derivation except in the sentential form $\alpha S \beta$. Reorder the derivation $\alpha S \beta \stackrel{*}{\Rightarrow} w$ into $\alpha S \beta \stackrel{*}{\Rightarrow} \gamma S \delta \stackrel{*}{\Rightarrow} w$ so that $\gamma, \delta \in \Sigma^*$. This implies $w = \gamma u \delta$ for some word $u \in L'$, where L' is defined as in Definition 12, completing the proof. ◀

4 The main results

Here we prove that the problem is decidable for two families of languages, namely the linear context-free languages and the linear bounded languages.

4.1 The bounded semilinear languages

We solve the easier case. We recall that a *bounded semilinear* language is a finite union of *linear* languages which are languages of the form

$$L = \{w_1^{n_1} \cdots w_k^{n_k} \mid (n_1, \ldots, n_k) \in R\} \tag{10}$$

for some fixed words $w_i \in \Sigma^*$ for $i = 1, \ldots, k$ and $R \subseteq \mathbb{N}^k$ is a linear set, i.e., there exists $v_0, v_1, \ldots, v_p \in \mathbb{N}^k$ such that

$$R = \{v_0 + \lambda_1 v_1 + \cdots + \lambda_p v_p \mid \lambda_1, \ldots, \lambda_p \in \mathbb{N}\}$$

▶ Proposition 8. If L is bounded semilinear then its closure $\mathbf{C}l(\varphi(L))$ is semialgebraic.

Furthermore, if the quantum automaton $\mathcal{Q}$ is rational, the $L - Q$-intersection is decidable.

For the proof of this proposition, see the Appendix.

4.2 The case of context-free languages

Here we show that $\mathbf{C}\mathrm{l}(\varphi(L))$ is effectively eventually definable for languages generated by linear grammars and rational quantum automata.

We adopt the notations of Section 3 for context-free grammars. We recall the following notion that will be used in the proof of the next result (see [8]). A subset of a monoid M is *regular* if it is recognized by some finite M-automaton which differs from an ordinary finite nondeterministic automaton over the free monoid by the fact the transitions are labeled by elements in M.

▶ Proposition 9. If L is generated by a context-free grammar, then $\mathbf{C}l(\varphi(L))$ is semialgebraic. Furthermore, if the grammar is linear and if the quantum automaton is rational then $\mathbf{C}l(\varphi(L))$ is effectively eventually definable and the $L - Q$-intersection is decidable.

Proof. With the notations of section 3 the language L is a finite union of languages of the form $C_S \diamond L''$ with

$$L'' = w_1 L_{A_1} w_2 L_{A_2} \cdots w_\ell L_{A_\ell} w_{\ell+1} \tag{11}$$

where, for every $1 \leq i \leq \ell + 1$, and $w_i \in \Sigma^*$ and $A_i \in V$. It suffices to show by induction on the number of nonterminal symbols that, with the previous notations, the subsets

$$\mathbf{C}l(\varphi(C_S \diamond L'')) \tag{12}$$

are semialgebraic in all cases and effectively eventually decidable when the quantum automaton is rational and the grammar of the language is linear. As a preliminary remark let us show this property for $\mathbf{C}l(\varphi(C_S))$. Define $\varphi^T : \Sigma^* \to \mathbb{R}^{n\times n}$ by posing $\varphi^T(u) = \varphi(u)^T$. Set

$$M = \{\varphi(a) \oplus \varphi^T(b^T)) \mid (a, b) \in C_S\}$$

where, applied to a matrix, the superscript T represents its transpose, while applied to a word, it represents its mirror image.

Observe that M is a submonoid since if $\varphi(a) \oplus \varphi^T(b)$ and $\varphi(c) \oplus \varphi^T(d)$ are in M then we have

$$\varphi^T(b)\varphi^T(d) = \varphi(b)^T\varphi(d)^T = (\varphi(d)\varphi(b))^T = \varphi(db)^T = \varphi^T(db)$$

which yields

$$(\varphi(a) \oplus \varphi^T(b))(\varphi(c) \oplus \varphi^T(d)) = \varphi(ac) \oplus \varphi^T(db)$$

Furthermore M is a regular submonoid of the group of orthogonal matrices $O_n \oplus O_n$ if the grammar is linear. Indeed, it is recognized by the finite O_{2n}-automaton whose states are the nonterminal symbols, the transitions are of the form $A \xrightarrow{(\varphi(a)\oplus\varphi^R(b)} B$ where $A \to aBb$ is a rule of the grammar and where the initial and final states coincide with S. As a first consequence, by Corollary 8, $\mathbf{C}l(M)$ is algebraic. Now, the subgroup generated by a regular subset of a monoid has an effective finite generating set, e.g., [1] (cf. also [8]) and thus by Proposition 2 $\mathbf{C}l(M)$ is effectively eventually definable if $\varphi(\Sigma^*) \subseteq \mathbb{Q}^{n\times n}$.

We proceed by induction on the number of nonterminals. If the set of nonterminals is reduced to S then L is reduced to L'. We may further assume that there is a unique terminal rule $S \to w$. By Theorem 1 we have

$$\mathbf{C}l(\varphi(L)) = \{X\varphi(w)Y^T \mid X \oplus Y \oplus \{\varphi(w)\} \in \mathbf{C}l(M \oplus \varphi(w))\}$$

By Corollary 2 we have

$$\mathbf{C}l(M \oplus \varphi(w)) = \mathbf{C}l(M) \oplus \mathbf{C}l(\varphi(w)) = \mathbf{C}l(M) \oplus \varphi(w)$$

which, by Proposition 6, is semialgebraic, resp. effectively eventually definable. In that latter case the $L - Q$-intersection is decidable.

Now assume V contains more than one variable. We first prove that for each nonterminal A, $\mathbf{C}l(\varphi(C_S \diamond L_A))$ is semialgebraic in the general case and effectively eventually definable

when the grammar is linear and the quantum automaton is rational. By Theorem 1 and Corollary 2, $\mathbf{C}l(\varphi(C_S \diamond L''))$ is the subset

$$\mathbf{C}l(\varphi(C_S \diamond L'')) = \{XZY^T \mid X \oplus Y \oplus Z\} \in \mathbf{C}l(M) \oplus \mathbf{C}l(\varphi(L''))\}$$

with L'' as in expression 11, i.e.,

$$\{XZY^T \mid X \oplus Y \oplus Z \in \mathbf{C}l(M) \oplus \mathbf{C}l(\varphi(w_1)\varphi(L_{A_1}) \cdots \varphi(w_\ell)\varphi(L_{A_\ell})\varphi(w_{\ell+1}))\}$$

By Cororally 2 we have

$$\begin{array}{rl} & \mathbf{C}l(\varphi(w_1)\varphi(L_{A_1}) \cdots \varphi(w_\ell)\varphi(L_{A_\ell})\varphi(w_{\ell+1}))\} \\ = & \varphi(w_1)\mathbf{C}l(\varphi(L_{A_1})) \cdots \varphi(w_\ell)\mathbf{C}l(\varphi(L_{A_\ell}))\varphi(w_{\ell+1})\} \end{array}$$

which shows, via Proposition 3 and by induction hypothesis that this subset is semialgebraic, resp. effectively, eventually decidable. Then its direct sum with $\mathbf{C}l(M)$ is semialgebraic and effectively, eventually decidable if the grammar is linear and the quantum automaton is rational. We conclude by applying Proposition 5. ◀

5 Complement of context-free languages

In this section we prove that there is a language L such that (i) the complement of L is linear context-free and (ii) $\mathbf{C}l(\varphi(L))$ is not semialgebraic.

Given a real $\alpha = 0.b_1 \cdots b_n \cdots$, we define its *approximation sequence* $(\alpha[k])_{k\geq 0}$ as the sequence of its successive truncations $\alpha[k] = 0.b_1 \dots b_k$.

▶ **Lemma 14.** *Let $0 < \alpha < 1$ be an irrational. There exist infinitely many rationals $\frac{q}{n}$ such that*

$$\left|\alpha[1 + 2\ell(n)] - \frac{q}{n}\right| < \frac{1}{n^2}$$

holds, where $\ell(n) = \lfloor \log_2 n \rfloor$.

Proof. By the triangular inequality we have

$$\left|\alpha[1 + 2\ell(n)] - \frac{q}{n}\right| \leq |\alpha[1 + 2\ell(n)] - \alpha| + \left|\alpha - \frac{q}{n}\right|$$

By the definition of the approximation sequence we get

$$|\alpha[1 + 2\ell(n)] - \alpha| \leq \frac{1}{2^{1+2\ell(n)}} = \frac{1}{2} \times \frac{1}{2^{2\lfloor \log_2 n \rfloor}} \leq \frac{1}{2} \times \frac{1}{n^2}$$

Now by Hurwitz Theorem there exist infinitely many rationals $\frac{p}{n}$ for which

$$\left|\alpha - \frac{p}{n}\right| \leq \frac{1}{\sqrt{5}} \times \frac{1}{n^2}$$

We conclude by combining these last two inequalities. ◀

We now fix an irrational $0 < \alpha < 1$. Consider the orthogonal matrix

$$M_\alpha = \begin{pmatrix} \cos 2\pi\alpha & \sin 2\pi\alpha \\ -\sin 2\pi\alpha & \cos 2\pi\alpha \end{pmatrix}$$

and the morphism $\varphi_\alpha : b^* \to O_n$ from the free monoid generated by the letter b and the group O_n defined by $\varphi_\alpha(b) = M_\alpha$. Furthermore set

$$L(\alpha) = \left\{ b^n \mid \exists\, q \in \mathbb{N} \mid \left|\alpha[1 + 2\ell(n)] - \frac{q}{n}\right| < \frac{1}{n^2} \right\} \tag{13}$$

▶ **Lemma 15.** *If $0 < \alpha < 1$ is an irrational, the topological closure $\mathbf{C}l(\varphi(L(\alpha))$ is not semialgebraic.*

Proof. Since the element in position $(1,1)$ of the matrix $\varphi_\alpha(b^n)$ is $\cos 2\pi n\alpha$ and since the projection of a semialgebraic set is semialgebraic, it suffices to show that

$$\mathbf{C}l(\{\cos 2\pi n\alpha \mid b^n \in L(\alpha)\})$$

is not semialgebraic.

Observe that $n \neq n'$ implies $\cos 2\pi n\alpha \neq \cos 2\pi n'\alpha$ since α is irrational. In particular, the set $\{\cos 2\pi n\alpha \mid b^n \in L(\alpha)\}$ is infinite.

Now we verify that 1 is the unique limit point. Indeed, by definition $b^n \in L(\alpha)$ implies that for some integer q we have

$$\left|\alpha[1+2\ell(n)] - \frac{q}{n}\right| \leq \frac{1}{n^2}$$

For such an integer q we have

$$\begin{aligned}\left|\alpha - \frac{q}{n}\right| &\leq |\alpha - \alpha[1+2\ell(n)]| + \left|\alpha[1+2\ell(n)] - \frac{q}{n}\right| \\ &\leq \frac{1}{2n^2} + \frac{1}{n^2} = \frac{3}{2} \times \frac{1}{n^2}\end{aligned}$$

Consequently, $|n\alpha - q| \leq \frac{3}{2} \times \frac{1}{n}$. Now we compute

$$1 \geq \cos 2\pi n\alpha = \cos 2\pi(n\alpha - q) = \cos 2\pi|n\alpha - q| \geq \cos\frac{3\pi}{n}$$

which proves that the closure $\mathbf{C}l(\{\cos 2\pi n\alpha \mid b^n \in L(\alpha)\})$ consists of a unique limit point and of infinitely many isolated points. This is not a semialgebraic set since the semialgebraic sets on the reals are finite unions of intervals. ◀

Now we state the main result of this section (for the proof of this theorem, see the Appendix).

▶ **Theorem 16.** *There is a language $L \subseteq \Sigma^*$ and a morphism $\varphi : \Sigma^* \to \mathbb{Q}^{n\times n}$, assigning an orthogonal matrix to every word of Σ^*, such that (i) L is the complement of a context-free language (ii) the topological closure $\mathbf{C}l(\varphi(L))$ is not semialgebraic.*

References

1. A. V. Anisimov, and F. D. Seifert. *Zur algebraischen Charakteristik der durch Kontext-freie Sprachen definierten Gruppen*. Elektron. Inform. Verarb. u. Kybern. 11, 695-702, 1975.
2. S. Basu, R. Pollack, and M. -F. Roy. *Algorithms in Real Algebraic Geometry*. Springer, Berlin, 2003.
3. V. D. Blondel, E. Jeandel, P. Koiran, and N. Portier. *Decidable and Undecidable Problems about Quantum Automata*. SIAM J. Comput. 34, 1464-1473, 2005.
4. H. Derksen, E. Jeandel, and P. Koiran. *Quantum automata and algebraic groups*. J. Symb. Comput. 39, 357-371, 2005.
5. J. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, 1979.
6. C. Moore, and J. Crutchfield. *Quantum automata and quantum grammars*. Theoret. Comput. Sci. 237, 275-306, 2000.
7. A. Onishchik and E. Vinberg. *Lie Groups and Algebraic Groups*. Springer, Berlin, 1990.
8. J. Sakarovitch. *Elements of Automata Theory*. Cambridge University Press, Cambridge, 2009.

A Appendix

Proof of Theorem 1: Since the function f is continuous, the inverse image of $\mathbf{C}l(f(\rho))$ is closed, i.,e., $\mathbf{C}l(\rho) \subseteq f^{-1}(\mathbf{C}l(f(\rho)))$ holds which yields $f(\mathbf{C}l(\rho)) \subseteq \mathbf{C}l(f(\rho))$. Now we prove the opposite inclusion. Consider an element $A \in \mathbf{C}l(f(\rho))$. It is the limit of a sequence $M_{1,n} \cdots M_{k,n}$ where $(M_{1,n}, \cdots, M_{k,n}) \in \rho$ for $n \geq 0$. Because $\mathcal{C}$ is a compact set, there exists a subsequence $(M_{1,n_i}, \cdots, M_{k,n_i}) \in \rho$, i.e., an infinite sequence of strictly increasing indices n_i which converges to a limit point $(A_1, \cdots, A_k) \in \mathbf{C}l(\rho)$. By continuity we have $f(A_1, \cdots, A_k) = A$ which shows that $\mathbf{C}l(f(\rho)) \subseteq f(\mathbf{C}l(\rho))$. ◀

Proof of Theorem 7: It is clear that $\mathbf{C}l(\langle E \rangle)$ is a subgroup of orthogonal matrices, say G. By [7, Thm 5, p. 133] this group is the zero set of all polynomials $p[x_{1,1}, \ldots, x_{n,n}]$ satisfying the conditions where I denotes the identity matrix

$$p(I) = 0 \quad \text{and } p(gX) = p(X) \quad \text{for all } g \in G \tag{14}$$

Let us verify that we may assume the above condition is satisfied by all $e \in E$. First, if it is the case, it is satisfied for all elements of the group $\langle E \rangle$. Now observe that condition (14) defines a linear constraint on the coefficients of the polynomial: if $V \in \mathbb{R}^d$ is the vector of coefficients of the polynomial p then the above equality can be expressed as a system of linear equations

$$MV = V$$

where the matrix M depends on g only, say $M = M_g$. Let

$$\lim_{i \to \infty} g_i = g \quad \text{and } M_{g_i} V = V \quad \text{for all } i \geq 0$$

Then by continuity we have $M_g V = V$.

The last assertion concerning the case where the coefficients are rational can be found in [3, Th. 3.1]. ◀

Proof of Proposition 8: Because the semialgebraic sets are closed under finite union, it suffices to consider the case where the language is of the form (10). For $t = 0, \ldots, p$ set $v_t^T = (v_{t,1}, \ldots, v_{t,k})$ and consider the orthogonal matrices

$$g_t = \begin{pmatrix} \varphi(w_1)^{v_{t,1}} & 0 & 0 & 0 \\ 0 & \varphi(w_2)^{v_{t,2}} & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \varphi(w_k)^{v_{t,k}} \end{pmatrix}$$

Set $G = \{g_i \mid i = 1, \ldots, p\}$. In virtue of Theorem 6 and Proposition 2 the set $\mathbf{C}l(G^*)$ is semialgebraic and it is effectively eventually definable if the coefficients of the quantum automata are rational. By Corollary 2 we have

$$\mathbf{C}l(g_0 G^*) = g_0 \mathbf{C}l(G^*)$$

and by Proposition 3 this product is semialgebraic (resp. and effectively eventually definable if the coefficents of the quantum automaton are rational). By Proposition 5, $\mathbf{C}l(\varphi(L))$ is semialgebraic (resp. and effectively eventually definable if the coefficients of the quantum automaton are rational). In the latter case the $L - Q$-intersection is decidable by Proposition 1 which completes the proof. ◀

Proof of Theorem 16: Consider a one tape Turing machine implementing the following procedure for recognizing the language $L(\alpha)$ defined in (13):

Input b^n
$A \leftarrow \alpha[1 + 2\ell(n)]$
$F \leftarrow 0$
for $q = 1$ to n, if $|A - \frac{q}{n}| < \frac{1}{n^2}$ then $F \leftarrow 1$
if $F = 1$ then write ab^n,
position the head on the rightmost occurrence of b,
change to a new state $\hat{q}$, move the reading head to the leftmost cell while staying in state $\hat{q}$
stop when reaching the occurrence a.

We know that the computation histories of a Turing machine, i.e., the set of sequences of configurations properly separated by a new symbol is, as a language, the intersection of two linear context-free languages (see, for instance, [5], Lemma 8.6). Let $\mathrm{Hist}(b^n)$ be the history associated to the input b^n. Let Γ be the disjoint union of the symbols comprising the input and tape alphabets along with the states including the special state $\hat{q}$. With $\alpha = \arctan \frac{3}{4}$ we get the orthogonal matrix

$$M_\alpha = \begin{pmatrix} \frac{3}{5} & \frac{4}{5} \\ -\frac{4}{5} & \frac{3}{5} \end{pmatrix}$$

Define the morphism $\varphi : \Gamma^* \to O_2$ by

$$\varphi(c) := \begin{cases} I & \text{if } c \in \Gamma \setminus \{\hat{q}\} \\ M_\alpha & \text{if } c = \hat{q} \end{cases}$$

By applying the result mentioned above to the Turing machine implementing the procedure for recognizing the language $L(\alpha)$, we have that there exist two linear context-free languages L_1 and L_2 such that

$$\mathrm{Hist}(L_\alpha) = L_1 \cap L_2 = (L_1^c \cup L_2^c)^c$$

Since L_1 and L_2 are linear and deterministic context-free, $L_1^c \cup L_2^c$ is linear (in general non-deterministic) context-free and thus $\mathrm{Hist}(L_\alpha)$ is the complement of a linear context-free language. But then

$$\varphi(\mathrm{Hist}(L_\alpha)) = \{M_\alpha^n \mid b^n \in L_\alpha\}$$

We conclude by applying Lemma 15. ◀