On the commutative equivalence of bounded context-free and regular languages: the semi-linear case

Flavio D'Alessandro
Dipartimento di Matematica,
Università di Roma "La Sapienza"
Piazzale Aldo Moro 2, 00185 Roma, Italy,
and

Department of Mathematics, Boğaziçi University
34342 Bebek, Istanbul, Turkey

Benedetto Intrigila
Dipartimento di Ingegneria dell'Impresa,
Università di Roma "Tor Vergata",
via del Politecnico, 1, 00133 Roma, Italy.

Abstract

This is the third paper of a group of three where we prove the following result. Let A be an alphabet of t letters and let $\psi : A^* \longrightarrow \mathbb{N}^t$ be the corresponding Parikh morphism. Given two languages $L_1, L_2 \subseteq A^*$, we say that L_1 is commutatively equivalent to L_2 if there exists a bijection $f : L_1 \longrightarrow L_2$ from L_1 onto L_2 such that, for every $u \in L_1$, $\psi(u) = \psi(f(u))$. Then every bounded context-free language is commutatively equivalent to a regular language.

Publisher version with DOI: https://doi.org/10.1016/j.tcs.2014.10.005

1 Introduction

Let $A = \{a_1, \ldots, a_t\}$ be an alphabet of t letters and let $\psi : A^* \longrightarrow \mathbb{N}^t$ be the corresponding Parikh morphism. Given two languages L_1 and L_2 over the alphabet A, we say that L_1 is *commutatively equivalent to* L_2 if there exists a bijection $f : L_1 \longrightarrow L_2$ from L_1 onto L_2 such that, for every $u \in L_1$, $\psi(u) = \psi(f(u))$. This is the third paper of a cycle of three (cf [6, 7]) where the proof of the following theorem is provided.

Theorem 1 *Every bounded semi-linear language L_1 is commutatively equivalent to a regular language L_2. Moreover the language L_2 can be effectively constructed starting from an effective presentation of L_1.*

As an immediate consequence of Theorem 1, we obtain the following result.

Theorem 2 *Every bounded context-free language L_1 is commutatively equivalent to a regular language L_2. Moreover the language L_2 can be effectively constructed starting from an effective presentation of L_1.*

It is worth noticing that Theorem 2 does not hold for an arbitrary context-free language. Indeed, if such a language would be commutatively equivalent to a regular one, then its generating function would be rational. On the other hand, there exist context-free languages whose generating functions are algebraic not rational, as for instance in the case of the Dyck languages, and, even, transcendental, as proven by Flajolet [9]. Theorem 2 naturally fits in the algebraic theory of bounded context-free languages developed by Ginsburg and Spanier. We refer to [6] for a more exhaustive description of the relationships between Theorem 2 and the above mentioned theory.

We would like to give a short description of some key aspects of the proof of Theorem 1 as well as of the relations of the present paper with the other cited ones [6, 7]. Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded semi-linear language where, for every $i = 1, \ldots, k$, u_i is a non-empty word over a given alphabet A. Let us consider the *Ginsburg map*

$$\varphi : \mathbb{N}^k \longrightarrow u_1^* \cdots u_k^*$$

which associates with every tuple $(\ell_1, \ldots, \ell_k)$ of non-negative integers, the word $\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}$. By a result of [11], there exists a semi-simple set B of $\mathbb{N}^k$ such that $\varphi(B) = L$ and φ is injective on B. We recall that B

admits a partition into a finite family of sets:

$$B = \bigcup_{i=0}^{n} B_i, \quad n \geq 1,$$

where B_0 is a finite set of vectors and, for every $i = 1, \ldots, n$, B_i is a simple set of dimension $k_i > 0$:

$$B_i = \mathbf{b}_0^{(i)} + \{\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}\}^{\oplus}, \tag{1}$$

where $\mathbf{b}_0^{(i)}, \mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$, are the vectors of the unambiguous representation of B_i. The proof of Theorem 1 is essentially based upon two main tools. The first one is a refinement of a technique developed in [6], while the second tool has been developed in [7]. The first tool has been conceived to prove the theorem under the assumption that, for every $i = 1, \ldots, n$ and for every $j = 1, \ldots, k_i$, the word $\varphi(\mathbf{b}_j^{(i)})$, that represents via φ the vector $\mathbf{b}_j^{(i)}$ of (1), contains at least two distinct letters.

The second tool provides the solution of Theorem 1 in the opposite case, that is, under the assumption that there exists a letter $a \in A$ such that, for every $i = 1, \ldots, n$, all the words $\varphi(\mathbf{b}_1^{(i)}), \ldots, \varphi(\mathbf{b}_{k_i}^{(i)})$ are powers of a. We treat such last case by reducing the study of commutative equivalence for languages to that of the *commutative equivalence for semi-linear sets of vectors*. More precisely, given two subsets S_1, S_2 of $\mathbb{N}^k$, we say that S_1 is commutatively equivalent to S_2 if there exists a bijection $f : S_1 \longrightarrow S_2$ from S_1 onto S_2 such that, for every $\mathbf{v} \in S_1$, $|\mathbf{v}| = |f(\mathbf{v})|$, where $|\mathbf{v}|$ denotes the sum of the components of $\mathbf{v}$. In [7] we prove that every semi-linear set of $\mathbb{N}^k$ is commutatively equivalent to a subset which is recognizable in $\mathbb{N}^k$ in the classical sense of Elgot and Mezei. As a straightforward consequence of the latter result, we derive the corresponding one for languages. By eventually combining the two techniques, we then provide the proof of Theorem 1.

We give the proof of Theorem 1 by considering the case of a binary alphabet A of two letters a and c. This special case allows us to simplify the exposition of the proof and, at the same time, shows all the essential aspects of the argument. In the Appendix, we will show how to adapt such proof to the case of a finite arbitrary alphabet.

The paper is structured as follows. In Section 2, some basic results about bounded semi-linear languages are introduced. In Section 3 we describe the geometrical decomposition of simple sets. In Section 4, the proof of Theorem

1 is presented. An Appendix reports some technical proofs and some examples.
Finally, we mention an open problem pointed out in [12]. Also in the theory of formal series there is a notion of commutative equivalence (see [1], Ch. 14). Given a $\mathbb{N}$-series $\sigma \in \mathbb{N}\langle\langle A\rangle\rangle$ over an alphabet A of non-commutative variables, the *commutative image* of σ is the $\mathbb{N}$-series $\psi(\sigma) \in \mathbb{N}[[A]]$ over the commutative variables $\psi(A)$ defined as: for every $u \in \psi(A^*)$, $(\psi(\sigma), u) = \sum_{\psi(w)=u} (\sigma, w)$. Given two series $\sigma_1, \sigma_2 \in \mathbb{N}\langle\langle A\rangle\rangle$, we say that σ_1 is *commutatively equivalent to* σ_2 if they have the same commutative image. In this context, as a possible extension of Theorem 2, one can ask whether every $\mathbb{N}$-algebraic series with bounded support is commutatively equivalent to a rational one. Theorem 2 seems to be a first step of the study of this problem.

2 Preliminaries

The aim of this section is to introduce some results concerning bounded semi-linear languages. We assume that the reader is familiar with the basic notions of this theory. The reader is referred to [10].

2.1 Basic notation

We find useful to recall the attention of the reader to some notation adopted in this paper.

The letter k is always used to denote the dimension of the underlying working monoid $\mathbb{N}^k$. A vector of $\mathbb{N}^k$ is denoted in **bold** as, for instance, for $\mathbf{v}$ which represents the vector $(v_1, \ldots, v_k)$, or $\mathbf{v}_j = (v_1^{(j)}, v_2^{(j)}, \ldots, v_k^{(j)})$. Moreover if the vector is indexed, as for instance for $\mathbf{v}_j$, its components are also denoted $(v_{j1}, v_{j2}, \ldots, v_{jk})$.

A set of vectors of $\mathbb{N}^k$ is always denoted by using capital letters like, for instance, X, Y, L, *etc.* Given a set S, a family of n pairwise disjoint sets $S_1, \ldots, S_n$, such that $S = \bigcup_{i=1}^n S_i$, is called a *decomposition* of S. The number n will be denoted by $\sharp(S)$.

2.2 Semi-linear sets of $\mathbb{N}^k$

The free abelian monoid on k generators is identified with $\mathbb{N}^k$ with the usual additive structure. Let $B = \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}$ be a finite subset of $\mathbb{N}^k$. Then we denote by $B^\oplus$ the submonoid of $\mathbb{N}^k$ generated by B, that is

$$B^\oplus \;=\; \mathbf{b}_1^\oplus + \cdots + \mathbf{b}_m^\oplus \;=\; \{n_1\mathbf{b}_1 + \cdots + n_m\mathbf{b}_m \;\mid\; n_i \in \mathbb{N},\ i = 1, \ldots, m\}.$$

In the sequel, in the formula above we will assume $B = \emptyset$ whenever $m = 0$.

Definition 1 *Let X be a subset of $\mathbb{N}^k$. Then*

1. *X is* linear *in $\mathbb{N}^k$ if $X = \mathbf{b}_0 + \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}^\oplus$, where $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ are vectors of $\mathbb{N}^k$,*
2. *X is* simple *in $\mathbb{N}^k$ if $\mathbf{b}_1, \mathbf{b}_2, \ldots, \mathbf{b}_m$ are linearly independent in $\mathbb{Q}^k$,*
3. *X is* semi-linear *in $\mathbb{N}^k$ if X is a finite union of linear sets in $\mathbb{N}^k$,*
4. *X is* semi-simple *in $\mathbb{N}^k$ if X is a finite disjoint union of simple sets.*

In the sequel, we will adopt the following terminology.

Convention If $X = \mathbf{b}_0 + \{\mathbf{b}_1, \ldots, \mathbf{b}_m\}^\oplus$ is a simple set, then the vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$, will be called the *(unambiguous) representation* of X. Moreover, $\mathbf{b}_0$ will be called the *constant vector* and $\mathbf{b}_1, \ldots, \mathbf{b}_m$ will be called the *generators* of the representation, respectively.

It is worth noticing that the uniqueness of the representation of a simple set is *folklore.* With every simple set B whose representation is given by the vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$, we can associate the number m called the *dimension* of B. One obviously has that $m \leq k$ and the dimension of a singleton is 0. By convention, the dimension of the empty set is -1. The following theorem by Eilenberg and Schützenberger [8] provides an important characterization of semi-linear sets.

Theorem 3 *Let X be a subset of $\mathbb{N}^k$. Then X is semi-linear in $\mathbb{N}^k$ if and only if X is semi-simple in $\mathbb{N}^k$.*

Theorem 3 is effective. Indeed, one can effectively represent a semi-linear set X as a semi-simple set. More precisely, one can effectively construct a finite family $\{V_i\}$ of finite sets of vectors such that the vectors in V_i form a representation of a simple set X_i and X is the disjoint union of the sets X_i.

The following proposition states a well-known result proved by Ginsburg and Spanier (see [10]).

Proposition 1 *The family of semi-linear sets of $\mathbb{N}^k$ is closed under the Boolean set operations.*

2.3 Bounded languages

Along all this paper, we let $A = \{a_1, \ldots, a_t\}$ be an alphabet of t letters and we let A^* be the free monoid generated by A. The empty word of A^* is denoted by 1_{A^*}. The length of every word u is denoted $|u|$. For every $a \in A$, the number of occurrences of a in u will be denoted $|u|_a$.

We let $\psi : A^* \longrightarrow \mathbb{N}^t$ be the Parikh map over A. Moreover, if $u_1, \ldots, u_k$ are k words of A^+, we let

$$\varphi : \mathbb{N}^k \longrightarrow u_1^* \cdots u_k^*, \tag{2}$$

be the *Ginsburg map* which associates with every tuple $(\ell_1, \ldots, \ell_k)$ of non-negative integers, the word $\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}$.

The following proposition provides a faithful description of a bounded semi-linear language ([11], see also [2, 3] for a proof).

Proposition 2 *Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded semi-linear language. Then there exists a semi-simple set B of $\mathbb{N}^k$ such that $\varphi(B) = L$ and φ is injective on B. Moreover, B can be effectively constructed.*

2.4 Some results of combinatorics on words

The content of this section has been already presented in [6]. To help the reader, we report it here *verbatim.* We recall some notions and elementary results of Combinatorics on words that are needed in this setting.

Definition 2 *Let L_1, L_2 be two languages over A. We say that L_1 is* commutatively equivalent to *L_2 if there exists a bijection $f : L_1 \longrightarrow L_2$ such that, for every $u \in L_1$, one has $\psi(u) = \psi(f(u))$.*

In the sequel, if L_1 is commutatively equivalent to L_2, we simply write $L_1 \sim L_2$. The following lemmas are easily proved.

Lemma 1 *Let L_1, L_2, L_1' and L_2' be languages over A. Suppose that $L_i \sim L_i'$ $(i = 1, 2)$ and $L_1 \cap L_2 = L_1' \cap L_2' = \emptyset$. Then $(L_1 \cup L_2) \sim (L_1' \cup L_2')$.*

Lemma 2 *Let L_1, L_2 be languages over A such that $L_1 \sim L_2$ and L_2 is regular. Let $w \in A^*$ such that $w \notin L_1$. Then there exists a regular language L_2' such that $\{w\} \cup L_1 \sim L_2'$.*

Proof. Set $[w] = \{u \in A^* : \psi(u) = \psi(w)\}$. Obviously, $[w] \not\subseteq L_1$. Since $L_1 \sim L_2$, the latter implies $[w] \not\subseteq L_2$, so that there exists a word w' such that $\psi(w) = \psi(w')$ and $w' \notin L_2$. Hence $L_1 \cup \{w\} \sim L_2'$, with $L_2' = L_2 \cup \{w'\}$. □

In the sequel,

$$u_1, \ldots, u_k,$$

will be a list of k non-empty words over the alphabet A, fixed once for all for the rest of the paper.

Lemma 3 *([6], Lemma 3) There exists a constant $\gamma \in \mathbb{N}$ such that the following condition holds: let $a, b \in A$, with $a \neq b$, and assume that w is a word having a factor of the form*

$$\underbrace{a^\gamma b a^\gamma b \ldots a^\gamma b a^\gamma}_{(k+1)-times} \tag{3}$$

where a^γ occurs $(k+1)$ times in the word (3). Then w is not a factor of any word in $u_1^* \cdots u_k^*$.

In the sequel, γ will denote the minimum constant specified by Lemma 3.

Let $\mathbf{v} = (v_1, \ldots, v_t) \in \mathbb{N}^t$ be a vector. We denote by $|\mathbf{v}|$ the non-negative integer $|\mathbf{v}| = v_1 + \cdots + v_t$. Let

$$\mathbf{w}_1, \ldots, \mathbf{w}_m, \tag{4}$$

be a list of (not necessarily pairwise distinct) m vectors of $\mathbb{N}^t$. We associate with it its corresponding multiset:

$$\{(\alpha_1, \mathbf{v}_1), \ldots, (\alpha_\ell, \mathbf{v}_\ell)\}, \tag{5}$$

where $\alpha_1 + \cdots + \alpha_\ell = m$ and, for every $i = 1, \ldots, \ell$, $\mathbf{v}_i$ is a vector of the list (4) and α_i is the number of vectors of (4) equals to $\mathbf{v}_i$.

Lemma 4 *([6], Lemma 4) Let us consider the list of vectors (4) together with its multiset (5). Suppose that:*

i) for every $j = 1, \ldots, m$, $\mathbf{w}_j$ has, at least, two non null components;

ii) for every $j = 1, \ldots, \ell$, $|\mathbf{v}_j| = \beta$, where β is a constant not depending on j.

Let N_j be the greatest integer such that $\mathbf{v}_j$ has the form $\mathbf{v}_j = N_j\bar{\mathbf{v}}_j$ with $\bar{\mathbf{v}}_j \in \mathbb{N}^t$, for every $j = 1, \ldots, \ell$. If, for every $j = 1, \ldots, \ell$,

$$N_j \geq m(\gamma + 1)(k + 1),$$

there exists a uniform code $\mathcal{W}$ of m (distinct) words of length β over the alphabet A such that

$$\forall\ i = 1, \ldots, \ell, \quad \mathrm{Card}(\{w \in \mathcal{W} \mid \psi(w) = \mathbf{v}_i\}) = \alpha_i. \tag{6}$$

Moreover every $w \in \mathcal{W}$ has a prefix of length $\gamma(k + 1) + k$ that cannot be a factor of any word in $u_1^* \cdots u_k^*$. In particular every $w \in \mathcal{W}$ is not a factor of any word in $u_1^* \cdots u_k^*$.

Remark 1 Roughly speaking, Lemma 4 states the following fact. We are given a distribution of Parikh vectors of words, where all of the words have the same length and contain at least two distinct letters. Under the assumption that all the components of every Parikh vector are sufficiently large, one can construct a uniform length code with the same distribution of Parikh vectors. Moreover, every word of the code is not a factor of any word of the set $u_1^* \cdots u_k^*$. This result will be used for the construction of a regular language which is commutatively equivalent to an arbitrarily given bounded semi-linear language contained in $u_1^* \cdots u_k^*$.

3 A geometrical decomposition of a simple set of $\mathbb{N}^k$

In this section, we introduce a slight refinement of a technique introduced in [6]. This technique of geometrical nature (inspired to our work [5]) provides a suitable decomposition of a simple set into the sets of integer points lying in the interiors of parallelepipeds of dimension lower than or equal to k. Let B be a simple set of $\mathbb{N}^k$ of dimension $m > 0$:

$$B = \mathbf{b}_0 + \mathbf{b}_1^{\oplus} + \cdots + \mathbf{b}_m^{\oplus} = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m \mid x_i \in \mathbb{N},\ i = 1, \ldots, n\},$$

where the vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of B. Let

$$(N_1 + \chi_1, \ldots, N_m + \chi_m) \tag{7}$$

be a tuple where, for every $i = 1, \ldots, m$, N_i and χ_i are given non-negative integers. Let $\{+, -\}$ be an alphabet of two symbols and let $\mathcal{E}$ be the set

$$\mathcal{E} = \{(\epsilon_1, \ldots, \epsilon_m) \ \mid \epsilon_i \in \{+, -\}, \ i = 1, \ldots, m\},$$

of all sequences of length m with elements in the set $\{+, -\}$. With every sequence $(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E}$, we associate the set

$$B^{(\epsilon_1, \ldots, \epsilon_m)} \tag{8}$$

given by all the vectors $\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m$, where, for every $i = 1, \ldots, m$, one has:

$$x_i \geq N_i + \chi_i \quad \text{if } \epsilon_i = +,$$

$$x_i < N_i + \chi_i \quad \text{if } \epsilon_i = -.$$

Observe that, for every $(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E}$, $B^{(\epsilon_1, \ldots, \epsilon_m)}$ is a semi-simple set. In order to simplify the notation, if, for every $i = 1, \ldots, m$, $\epsilon_i = -$, then the corresponding set $B^{(\epsilon_1, \ldots, \epsilon_m)}$ will be simply denoted B^-. Observe that B^- is the unique finite set of the family (8). From the fact that the vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of the simple set B, we have:

Lemma 5 *The family $\{B^{(\epsilon_1, \ldots, \epsilon_m)}\}_{(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E}}$ gives a partition of B.*

Let $(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E} \setminus \{(-, -, \ldots, -)\}$, that is, there exists i, with $1 \leq i \leq m$, where $\epsilon_i = +$. Then there exists a non-negative integer p, depending on $(\epsilon_1, \ldots, \epsilon_m)$, such that the set of indices i, with $i = 1, \ldots, m$ is partitioned in two sets:

$$I^-_{\epsilon_1 \cdots \epsilon_m} = \{i_1, \ldots, i_p\}, \quad I^+_{\epsilon_1 \cdots \epsilon_m} = \{i_{p+1}, \ldots, i_m\}, \tag{9}$$

where

$$\epsilon_{i_\ell} = -, \ \ell = 1, \ldots, p, \quad \epsilon_{i_\ell} = +, \ \ell = p+1, \ldots, m.$$

It's worth to remak that:

- If, for every $i = 1, \ldots, m$, $\epsilon_i = +$, then one has $I^-_{\epsilon_1 \cdots \epsilon_m} = \emptyset$;
- the integer p depends upon the sequence $(\epsilon_1, \ldots, \epsilon_m)$; in the sequel, if no ambiguity arises, we will drop the dependency of p from the sequence $(\epsilon_1, \ldots, \epsilon_m)$.

Let us now associate with every index i_ℓ of the set $I^+_{\epsilon_1 \cdots \epsilon_m}$ of (9) a remainder r_{i_ℓ} modulo N_{i_ℓ}, where all the constants N_{i_ℓ} are defined in (7).

Similarly, let us associate with every index i_ℓ of the set $I^-_{\epsilon_1 \cdots \epsilon_m}$ of (9) a non-negative integer $c_{i_\ell} < N_{i_\ell} + \chi_{i_\ell}$, with respect to the constants defined in (7). We thus obtain a sequence of m constants

$$(c_{i_1}, \ldots, c_{i_p}, r_{i_{p+1}}, \ldots, r_{i_m}), \tag{10}$$

Denote by $\mathcal{C}_{\epsilon_1 \cdots \epsilon_m}$ the set of all sequences (10).

For every sequence $(c_{i_1}, \ldots, c_{i_p}, r_{i_{p+1}}, \ldots, r_{i_m}) \in \mathcal{C}_{\epsilon_1 \cdots \epsilon_m}$, define the simple set of vectors $B(\epsilon_1, \ldots, \epsilon_m, c_{i_1}, \ldots, c_{i_p}, r_{i_{p+1}}, \ldots, r_{i_m})$ as:

$$\{\mathbf{b}_0 + \sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} (r_{i_\ell} + \chi_{i_\ell})\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} N_{i_\ell} x_{i_\ell} \mathbf{b}_{i_\ell} \mid x_{i_\ell} \geq 1\}. \tag{11}$$

In the sequel, to shorten the notation, we denote a set of the family (11) as:

$$B(\epsilon_1, \ldots, \epsilon_m, d_1, \ldots, d_m),$$

where it is understood that the sequence of numbers $(d_1, \ldots, d_m)$ is defined as:

$$d_1 = c_{i_1}, \ldots, d_p = c_{i_p}, \quad d_{p+1} = r_{i_{p+1}} + \chi_{i_{p+1}}, \ldots, \; d_m = r_{i_m} + \chi_{i_m}.$$

Proposition 3 *The sets (11), together with the set B^-, give a partition of B into a finite union of pairwise disjoint semi-simple sets.*

Proof. The claim follows immediately from (11) and from the fact that the vectors $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of B. □

An application of Proposition 3 is shown in Example 1 of the Appendix.

Let us associate with every set $B(\epsilon_1, \ldots, \epsilon_m, d_1, \ldots, d_m)$ of (11) the language of A^*:

$$L(\epsilon_1, \ldots, \epsilon_m, d_1, \ldots, d_m) = \varphi(B(\epsilon_1, \ldots, \epsilon_m, d_1, \ldots, d_m)), \tag{12}$$

and let $L^- = \varphi(B^-)$. The following result follows immediately from Proposition 2 and Proposition 3.

Proposition 4 *The family (12), together with L^-, gives a partition of L.*

4 The construction of the regular language

In this section, we prove Theorem 1. Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded semi-linear language and let $\varphi : \mathbb{N}^k \longrightarrow u_1^* \cdots u_k^*$ be the map defined in (2). By Proposition 2 there exists a semi-simple set B of $\mathbb{N}^k$ such that $\varphi(B) = L$ and φ is injective on B. In the sequel, we will assume that B is infinite since, otherwise, Theorem 1 is trivially proved. We give the proof of Theorem 1 by considering the case of a binary alphabet A of two letters a and c. This assumption allows us to strongly simplify the exposition of the proof and, at same time, shows all the essential aspects of the arguments. In the Appendix, we will show how to adapt such proof to the case of a finite arbitrary alphabet.

From now on, we will suppose that L is a bounded semi-linear language such that $L = \varphi(B)$ and B is partitioned into a finite family of pairwise disjoint sets

$$B = B_0 \cup B_1 \cup \cdots \cup B_n,$$

where B_0 is finite and, for every $i = 1, \ldots, n$, B_i is a simple set of dimension $k_i > 0$

$$B_i = \mathbf{b}_0^{(i)} + \{\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}\}^{\oplus}, \tag{13}$$

where $\mathbf{b}_0^{(i)}$, $\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$ is the representation of B_i. The combinatorial structure of the words that represent, *via* φ, the generators of the simple sets B_i of (13) plays a crucial role in our solution. For this reason, we find useful to adopt the formalism described in the next section.

4.1 An enumeration of the generators of the set B_i

Let B_i, with $i = 1, \ldots, n$, be a simple set of the decomposition (13) of B. Let $\mathbf{b}_0^{(i)}$, $\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$ be the vectors of the representation of B_i, where $k_i > 0$ is the dimension of B_i.

From now on, we will fix, once for all, an enumeration of the generators of B_i such that three integers $n_a^{(i)}$, $n_c^{(i)}$, $n_+^{(i)} \in \mathbb{N}$, are determined according to the following property:

1. $\forall\ \ell = 1, \ldots, n_a^{(i)}$, $\varphi(\mathbf{b}_\ell^{(i)}) \in a^+$,

2. $\forall\ \ell = 1, \ldots, n_c^{(i)}$, $\varphi(\mathbf{b}_\ell^{(i)}) \in c^+$,

3. $\forall\ \ell = 1, \ldots, n_+^{(i)}$, $\varphi(\mathbf{b}_\ell^{(i)})$ contains at least two distinct letters.

If no ambiguity arises, in the sequel, we will denote $n_a^{(i)}$, $n_c^{(i)}$, $n_+^{(i)}$ by n_a, n_c, n_+ respectively, dropping the dependency of such numbers from the index i. Therefore, we can display the generators of B_i as:

$$\mathbf{b}_{a,1}^{(i)}, \ldots, \mathbf{b}_{a,n_a}^{(i)},\ \mathbf{b}_{c,1}^{(i)}, \ldots, \mathbf{b}_{c,n_c}^{(i)},\ \mathbf{b}_{+,1}^{(i)}, \ldots, \mathbf{b}_{+,n_+}^{(i)}, \tag{14}$$

where it is understood that:

1. $\forall\ \ell = 1, \ldots, n_a,\ \varphi(\mathbf{b}_{a,\ell}^{(i)}) \in a^+$,
2. $\forall\ \ell = 1, \ldots, n_c,\ \varphi(\mathbf{b}_{c,\ell}^{(i)}) \in c^+$,
3. $\forall\ \ell = 1, \ldots, n_+,\ \varphi(\mathbf{b}_{+,\ell}^{(i)})$ contains at least two distinct letters.

4.2 Determining the constants $N_j^{(i)}$

The aim of this section is the following. For every simple set B_i of (13), we want to compute the geometrical decomposition defined in Section 3. For this purpose, we need to determine a suitable sequence of constants of type (7). Let c be a non-negative integer and let $\beta(c)$ be the positive integer defined as

$$\beta(c) = \Pi_{i=1}^{n}\ \Pi_{j=1}^{k_i}\ |\varphi(\mathbf{b}_j^{(i)})|c,$$

where, for every $i = 1, \ldots, n$, and for every $j = 1, \ldots, k_i$, $\mathbf{b}_j^{(i)}$ is the j-th generators of the simple set B_i of (13). For every $i = 1, \ldots, n$, and for every $j = 1, \ldots, k_i$, let $N_j^{(i)}(c)$ be the number defined as:

$$N_j^{(i)}(c) \ = \ \left(\frac{\beta(c)}{|\varphi(\mathbf{b}_j^{(i)})|} \right). \tag{15}$$

The following lemma is immediate.

Lemma 6 *Let $N_j^{(i)}(c)$ be the numbers defined in (15). For every $i = 1, \ldots, n$, and for every $j = 1, \ldots, k_i$, one has $|\varphi(N_j^{(i)}(c)\mathbf{b}_j^{(i)})| = \beta(c)$.*

From now on, in all the rest of the paper, we will assume that c is the minimum positive integer such that, for every $i = 1, \ldots, n$, and for every $j = 1, \ldots, k_i$:

$$N_j^{(i)}(c) \geq m(\gamma + 1)(k + 1), \tag{16}$$

where $m = k_1 + \cdots + k_n$ is the sum of the dimensions of the simple sets B_i and γ is the fixed constant of Lemma 3. By the sake of simplicity, from now on, for every $i = 1, \dots, n$, the above defined numbers $N_j^{(i)}(c)$ will be denoted

$$N_1^{(i)}, N_2^{(i)}, \dots, N_{k_i}^{(i)}, \tag{17}$$

and the corresponding number $\beta(c)$ will be denoted β. Moreover, for every $i = 1, \dots, n$, and for every $j = 1, \dots, k_i$, we set

$$\chi_j^{(i)} = N_j^{(i)}(k_i + 1)\beta. \tag{18}$$

4.3 A coding

Let us describe first the aim of this section. For every simple set B_i, with $i = 1, \dots, n$, of the decomposition (13), consider the generators whose image, via the map φ, are words that contain at least two distinct letters. We want to codify such vectors with words of a uniform length code.

For this purpose, let B_i, with $i = 1, \dots, n$, be a simple set of the decomposition (13) where $\mathbf{b}_0^{(i)},\ \mathbf{b}_1^{(i)}, \dots, \mathbf{b}_{k_i}^{(i)}$ is the representation of B_i. Let us consider the (possibly empty) set of vectors

$$N_{+,1}^{(i)}\mathbf{b}_{+,1}^{(i)}, \dots, N_{+,n_+}^{(i)}\mathbf{b}_{+,n_+}^{(i)}, \tag{19}$$

where, according to the enumeration (14), for every $\ell = 1, \dots, n_+$, $\varphi(\mathbf{b}_{+,\ell}^{(i)})$ contains at least two distinct letters, and $N_{+,\ell}^{(i)}$ denotes the coefficient (17) associated with $\mathbf{b}_{+,\ell}^{(i)}$. Then, if the list above is not empty, we associate with B_i the list of (not necessarily pairwise distinct) words

$$\varphi(N_{+,1}^{(i)}\mathbf{b}_{+,1}^{(i)}),\ \dots,\ \varphi(N_{+,n_+}^{(i)}\mathbf{b}_{+,n_+}^{(i)}),$$

and the list of the corresponding Parikh vectors

$$\psi(\varphi(N_{+,1}^{(i)}\mathbf{b}_{+,1}^{(i)})),\ \dots,\ \psi(\varphi(N_{+,n_+}^{(i)}\mathbf{b}_{+,n_+}^{(i)})). \tag{20}$$

By performing the latter operation with every simple set B_i with $i = 1, \dots, n$ we construct a list of (not necessarily pairwise distinct) h Parikh vectors with $h \leq m = k_1 + \cdots + k_n$:

$$\mathbf{w}_1, \dots, \mathbf{w}_h. \tag{21}$$

Now by (16), the list of Parikh vectors (21) satisfies the hypotheses of Lemma 4 and thus, by applying this lemma, one gets the following result.

Lemma 7 *There exists a code $\mathcal{W} = \{w_1, \dots, w_h\}$ of h distinct words of length β such that, for every $j = 1, \dots, h$, $\psi(w_j) = \mathbf{w}_j$, that is, the Parikh vector of w_j is the vector $\mathbf{w}_j$ of the list (21). Moreover every $w \in \mathcal{W}$ has a prefix of length $\gamma(k+1)+k$ that cannot be a factor of any word in $u_1^* \cdots u_k^*$. In particular every $w \in \mathcal{W}$ is not a factor of any word in $u_1^* \cdots u_k^*$.*

Remark 2 Let w_j be a word of the code $\mathcal{W}$.

(i) w_j has a prefix of length $\gamma(k+1)+k$ that cannot be a factor of any word in $u_1^* \cdots u_k^*$. In particular every w_j is not a factor of any word in $u_1^* \cdots u_k^*$.

(ii) $|w_j| = \beta$.

By the previous lemma, there exists a uniform length code of h distinct words

$$w_1, \dots, w_h, \tag{22}$$

where, for every $j = 1, \dots, h$, $\psi(w_j) = \mathbf{w}_j$. Finally we define our desired coding as the one-to-one correspondence:

$$N_{+,\ell}^{(i)} \mathbf{b}_{+,\ell}^{(i)} \longrightarrow w_{+,\ell} \tag{23}$$

that, given a simple set B_i, with $1 \leq i \leq n$, of the decomposition (13), maps every vector $N_{+,\ell}^{(i)} \mathbf{b}_{+,\ell}^{(i)}$ of the list (19) into exactly one code word $w_{+,\ell}$ of the list (22) in such a way that $\psi(\varphi(N_{+,\ell}^{(i)} \mathbf{b}_{+,\ell}^{(i)})) = \psi(w_{+,\ell})$.

4.4 Normalization

The aim of this section is the following. By applying the technique of Section 3, we want to decompose the set B according to the combinatorial structure of the words that represent, *via* the Ginsburg map φ, the generators of the simple sets B_i of the decomposition (13).

Let B_i, with $1 \leq i \leq n$, be a simple set of the decomposition (13) and let us consider the sequence of integers (17):

$$N_1^{(i)}, N_2^{(i)}, \dots, N_{k_i}^{(i)},$$

together with the sequence (18):

$$\chi_1^{(i)} = N_1^{(i)}(k_i+1)\beta,\ \chi_2^{(i)} = N_2^{(i)}(k_i+1)\beta,\ \dots,\ \chi_{k_i}^{(i)} = N_{k_i}^{(i)}(k_i+1)\beta.$$

By Proposition 3, starting from the two sequences above, there exists a finite family of simple sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined by Eq. (11) such that

$$B_i = B_i^- \cup \bigcup B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}),$$

where $B_i^- = B_i^{(\epsilon_1, \ldots, \epsilon_{k_i})}$ is the set associated with the sequence $(-, -, \ldots, -)$.

Remark 3 The bold number $\mathbf{i}$ that appears in $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ is used to emphasize the fact that such set belongs to the partition (11) of B_i.

Hence we have:

$$B = B_0 \ \cup \bigcup_{i=1}^{n} \ (\bigcup B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})).$$

Now we simply re-arrange the sets of the partition above of B as:

$$B = B_- \cup B_+ \cup B_{a,c} \cup B_a \cup B_c, \tag{24}$$

where the semi-simple sets $B_-, B_+, B_{a,c}, B_a, B_c$ are defined as follows:

1. $B_- = B_0 \ \cup \ \bigcup_{i=1}^{n} B_i^-$;

2. B_+ is the union of all the sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following property: there exists ℓ with $1 \leq \ell \leq k_i$ where $\epsilon_\ell = +$ and $\varphi(N_\ell^{(i)} \mathbf{b}_\ell^{(i)})$ contains at least two distinct letters;

3. $B_{a,c}$ is the union of all the sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following properties:

- **3.1.** there exist two indices ℓ_a and ℓ_c with $1 \leq \ell_a, \ell_c \leq k_i$ such that $\epsilon_{\ell_a} = \epsilon_{\ell_c} = +$ and $\varphi(N_{\ell_a}^{(i)} \mathbf{b}_{\ell_a}^{(i)}) \in a^+$, $\varphi(N_{\ell_a}^{(i)} \mathbf{b}_{\ell_c}^{(i)}) \in c^+$;
- **3.2.** no index ℓ exists such that $\epsilon_\ell = +$ and $\varphi(N_\ell^{(i)} \mathbf{b}_\ell^{(i)})$ contains at least two distinct letters;

4. B_a is the union of all the sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following property: for every $\ell = 1, \ldots, k_i$, such that $\epsilon_\ell = +$, then $\varphi(\mathbf{b}_\ell^{(i)}) \in a^+$.

5. B_c is the union of all the sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following property: for every $\ell = 1, \ldots, k_i$, such that $\epsilon_\ell = +$, then $\varphi(\mathbf{b}_\ell^{(i)}) \in c^+$.

Let us set

$$L^- = \varphi(B^-),\ L_+ = \varphi(B_+),\ L_{a,c} = \varphi(B_{a,c}),\ L_a = \varphi(B_a),\ L_c = \varphi(B_c). \quad (25)$$

The following result is an immediate corollary of Proposition 2.

Corollary 1 *The languages of the family (25) gives a partition of L.*

Let

$$N = \max\{N_j^{(i)} + \chi_j^{(i)}\ :\ 1 \le i \le n,\ 1 \le j \le k_i\},$$

be the maximum of the integers defined, for every $i = 1, \ldots, n$, by (17) together with (18). If $\sigma \in A$ and $u \in A^*$, denote by $|u|_\sigma$ the number of occurrences of the symbol σ in u. Let $\mathrm{Max_c}$ be the number defined as

$$\mathrm{Max_c} = (1 + kN) \max\{|\varphi(\mathbf{b}_j^{(i)})|_c : 1 \le i \le n,\ 0 \le j \le k_i\}. \quad (26)$$

From the definition of B_a, it follows that

$$\forall\ u \in L_a, |u|_c < \mathrm{Max_c}\,. \quad (27)$$

Similarly, if we let $\mathrm{Max_a}$ be the number defined as

$$\mathrm{Max_a} = (1 + kN) \max\{|\varphi(\mathbf{b}_j^{(i)})|_a : 1 \le i \le n, 0 \le j \le k_i\}, \quad (28)$$

from the definition of B_c, it follows that

$$\forall\ u \in L_c, |u|_a < \mathrm{Max_a}\,. \quad (29)$$

We finally show that, up to a slight refinement of the decomposition (24) of the semi-simple B, we can assume that the following property holds

$$\forall\ u \in L_c,\ \mathrm{Max}_c \le |u|_c. \quad (30)$$

For this purpose, we describe an algorithm that provides the desired decomposition of B. Let $M = \mathrm{Max_c}$. Let $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be an arbitrary simple set of the decomposition of B_c and denote it by $\mathcal{B}$. According to (11) and (18), $\mathcal{B}$ can be written as

$$\mathcal{B} = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m : x_1, \ldots, x_m \ge 2\beta\}, \quad (31)$$

where $\mathbf{b}_0, \ldots, \mathbf{b}_m$ are vectors of $\mathbb{N}^k$, with $m \ge 1$, such that $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are linearly independent and, for every $\ell = 1, \ldots, m$, $\varphi(\mathbf{b}_\ell) \in c^+$.

In order to prove (30), it is sufficient to show that, for every set $\mathcal{B}$ of type (31), one has $|\varphi(\mathbf{b}_0)|_c \geq M$. Suppose that, for some of such set $\mathcal{B}$, $|\varphi(\mathbf{b}_0)|_c < M$. We can write $\mathcal{B}$ as $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$, with

– $\mathcal{B}_1 = \{\mathbf{b}_0 + \mathbf{b}_m + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m : x_1, \ldots, x_m \geq 2\beta\}$,
– $\mathcal{B}_2 = \{\mathbf{b}_0 + 2\beta\mathbf{b}_m + x_1\mathbf{b}_1 + \cdots + x_{m-1}\mathbf{b}_{m-1} : x_1, \ldots, x_{m-1} \geq 2\beta\}$.

Observe now that, for every $u \in \varphi(\mathcal{B}_1) \cup \varphi(\mathcal{B}_2)$, $|\varphi(\mathbf{b}_0)|_c + |\varphi(\mathbf{b}_m)|_c \leq |u|_c$. By iterating finitely many times the latter argument, one yields a decomposition of $\mathcal{B}$ as $\mathcal{B} = (\mathcal{B})_{\geq M} \cup (\mathcal{B})_{<M}$, where:

– $(\mathcal{B})_{\geq M}$ is a finite union of pairwise disjoint simple sets, every one of which is still in the form (31) and, for every $u \in \varphi((\mathcal{B})_{\geq M})$, $M \leq |u|_c$;
– $(\mathcal{B})_{<M}$ is a finite set of vectors.

Replace $\mathcal{B}$ with $(\mathcal{B})_{\geq M}$ and add (set-theoretically) $(\mathcal{B})_{<M}$ to the set B_- of the decomposition (24). By applying the argument above to every set $\mathcal{B}$ of the decomposition of B_c, one yields two new sets $(B_c)^{(new)}$ and $(B_-)^{(new)}$, where $\varphi((B_c)^{(new)})$ satisfies (30) and $(B_-)^{(new)}$ is finite. The required new decomposition of B is finally obtained by replacing, in the decomposition (24) of B, B_c with $(B_c)^{(new)}$ and B_- with $(B_-)^{(new)}$.

An immediate consequence of Eq. (27) and Eq. (30) is the following result.

Corollary 2 *Let L'_1 and L'_2 be languages such that $L'_1 \sim L_a$ and $L'_2 \sim L_c$. Then the languages L'_1 and L'_2 are disjoint.*

4.5 The clusterization of L_+

In this section, we will construct a regular language L'_+ such that L'_+ is commutatively equivalent to L_+. We call such construction *clusterization.* We will essentially use a refinement of the technique used in [6].

4.5.1 The definition of L'_+

We recall that, according to point (2) of (24), $L_+ = \varphi(B_+)$, where B_+ is the union of all the simple sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following property: there exists ℓ with $1 \leq \ell \leq k_i$ where $\epsilon_\ell = +$ and $\varphi(N_\ell^{(i)}\mathbf{b}_\ell^{(i)})$ contains at least two distinct letters.

Let $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be a simple set that appears in the decomposition of B_+. For the sake of simplicity, denote it by $\mathcal{B}$. We want now to associate with $\mathcal{B}$ a regular language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$. For this purpose, it is useful to remind the definition of $\mathcal{B}$. By (11), (17), and (18), the set $\mathcal{B}$ has the form:

$$\mathcal{B} = \{\mathbf{b}_0^{(i)} \ + \ \sum_{\ell=1}^{k_i} d_\ell \mathbf{b}_\ell^{(i)} \ + \ \sum_{\ell=1}^{s} x_{i_\ell} N_{i_\ell}^{(i)} \mathbf{b}_{i_\ell}^{(i)} \ \mid \ x_{i_\ell} \geq 1\}, \tag{32}$$

where:

- the vectors $\mathbf{b}_0^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$ that appear in (32) form the representation of B_i and k_i is the dimension of B_i; moreover $s \geq 1$;
- there exists at least an index $\ell = 1, \ldots, k_i$, such that $d_{i_\ell} \geq N_{i_\ell}^{(i)}(k_i+1)\beta$.

For the sake of simplicity, from now on, we will suppose that $i_\ell = i_1$.

According to the enumeration (14) of the generators of B_i defined in Section 4.1, we can display the generators $\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$of B_i as:

$$\mathbf{b}_{a,1}^{(i)}, \ldots, \mathbf{b}_{a,n_a}^{(i)}, \tag{33}$$

$$\mathbf{b}_{c,1}^{(i)}, \ldots, \mathbf{b}_{c,n_c}^{(i)}, \tag{34}$$

$$\mathbf{b}_{+,1}^{(i)}, \ldots, \mathbf{b}_{+,n_+}^{(i)}, \tag{35}$$

where $n_a, n_c, n_+ \in \mathbb{N}$ and

- $\forall \ \ell = 1, \ldots, n_+, \ \varphi(\mathbf{b}_{+,\ell}^{(i)})$ contains at least two distinct letters;
- $\forall \ \ell = 1, \ldots, n_a, \ \varphi(\mathbf{b}_{a,\ell}^{(i)}) \in a^+$,
- $\forall \ \ell = 1, \ldots, n_c, \ \varphi(\mathbf{b}_{c,\ell}^{(i)}) \in c^+$.

Thus, by (33), (34), and (35), the formal sum of vectors $\sum_{\ell=1}^{s} \ x_{i_\ell} N_{i_\ell}^{(i)} \mathbf{b}_{i_\ell}^{(i)}$ that appears in the expression (32) can be rewritten as:

$$\sum_{\ell=1}^{p_a} x_{a,i_\ell} N_{a,i_\ell}^{(i)} \mathbf{b}_{a,i_\ell}^{(i)} \ + \ \sum_{\ell=1}^{p_c} y_{c,i_\ell} N_{c,i_\ell}^{(i)} \mathbf{b}_{c,i_\ell}^{(i)} \ + \ \sum_{\ell=1}^{p_+} z_{+,i_\ell} N_{+,i_\ell}^{(i)} \mathbf{b}_{+,i_\ell}^{(i)}, \tag{36}$$

where we have:

– $0 \leq p_a \leq n_a, \ 0 \leq p_c \leq n_c, \ 1 \leq p_+ \leq n_+$,

– $N^{(i)}_{a,\ell}$ (resp., $N^{(i)}_{c,\ell}$, $N^{(i)}_{+,\ell}$) denotes the coefficient (17) associated with $\mathbf{b}^{(i)}_{a,\ell}$ (resp., $\mathbf{b}^{(i)}_{c,\ell}$, $\mathbf{b}^{(i)}_{+,\ell}$) in the expression (32),

– x_{a,i_ℓ}, y_{c,i_ℓ} and z_{+,i_ℓ} are free variables over $\mathbb{N}_+$. In the sequel, to simplify the notation, such variables will be denoted by x_{i_ℓ}, y_{i_ℓ} and z_{i_ℓ}, respectively.

Remark 4 Since we are dealing with the set B_+, one has n_+, $p_+ \geq 1$. Moreover, it is understood that, if $p_a = 0$ or $p_c = 0$, the corresponding sum in the expression (36) vanishes.

By (36), the set $\mathcal{B}$ can be rewritten as:

$$\{\mathbf{b}^{(i)} + \sum_{\ell=1}^{p_a} x_{i_\ell} N_{a,i_\ell} \mathbf{b}^{(i)}_{a,i_\ell} + \sum_{\ell=1}^{p_c} y_{i_\ell} N_{c,i_\ell} \mathbf{b}^{(i)}_{c,i_\ell} + \sum_{\ell=1}^{p_+} z_{i_\ell} N_{+,i_\ell} \mathbf{b}^{(i)}_{+,i_\ell}, \mid x_{i_\ell}, y_{i_\ell}, z_{i_\ell} \geq 1\}. \tag{37}$$

with $\mathbf{b}^{(i)} = \mathbf{b}^{(i)}_0 + \sum_{\ell=1}^{k_i} d_\ell \mathbf{b}^{(i)}_\ell$. By using the coding (23) of Section 4.3, we associate bijectively with every vector $N^{(i)}_{+,i_\ell} \mathbf{b}^{(i)}_{+,i_\ell}$ of the list

$$N^{(i)}_{+,i_1} \mathbf{b}^{(i)}_{+,i_1}, N^{(i)}_{+,i_2} \mathbf{b}^{(i)}_{+,i_2}, \ldots, N^{(i)}_{+,i_{p_+}} \mathbf{b}^{(i)}_{i_{p_+}},$$

a unique code word of the list

$$w_{i_1}, w_{i_2}, \ldots, w_{i_{p_+}}, \tag{38}$$

where such words are defined in (22). Now let us define the language $\mathrm{L}^{(i)}_a$ as:

$$\mathrm{L}^{(i)}_a = w_{i_1} L_{a,1} w_{i_1} L_{a,2} w_{i_1} \cdots w_{i_1} L_{a,n_a},$$

where:

- n_a is the length of the list (33);
- w_{i_1} is the first codeword of (38) and it occurs n_a times in $\mathrm{L}^{(i)}_a$;
- For every $j = 1, \ldots, n_a$, we set $L_{a,j} = (a^\beta)^+$ if the vector $\mathbf{b}^{(i)}_{a,j}$ of the list (33) appears in the sum (36); otherwise it is equal to 1_{A^*}.

Similarly, let us define the language $\mathrm{L}^{(i)}_c$ as:

$$\mathrm{L}^{(i)}_c = w_{i_1} L_{c,1} w_{i_1} L_{c,2} w_{i_1} \cdots w_{i_1} L_{c,n_c},$$

where:

- n_c is the length of the list (34);
- w_{i_1} is the first codeword of (38) and it occurs n_c times in $\mathrm{L}_c^{(i)}$;
- For every $j = 1, \ldots, n_c$, we set $L_{c,j} = (c^\beta)^+$ if the vector $\mathbf{b}_{c,j}^{(i)}$ of the list (34) appears in the sum (36); otherwise it is equal to 1_{A^*}.

Remark 5 It is understood that if $n_a = 0$ (resp. $n_c = 0$), $\mathrm{L}_a^{(i)}$ (resp. $\mathrm{L}_c^{(i)}$) vanishes. The interest of the language $\mathrm{L}_a^{(i)}$ relies on the following argument: every vector $N_{a,i_\ell}^{(i)}\mathbf{b}_{a,i_\ell}^{(i)}$ of $\mathcal{B}$ is represented univocally by the presence into $\mathrm{L}_a^{(i)}$ of the factor $L_{a,i_\ell} = (a^\beta)^+$ in the corresponding *slot* i_ℓ. The same remark holds for $\mathrm{L}_c^{(i)}$.

Finally we associate with $\mathcal{B}$ the regular language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined as:

$$\varphi(\widetilde{\mathbf{b}}^{(i)}) w_{i_1}^{(k_i - n_a - n_c)} \mathrm{L}_a^{(i)} \mathrm{L}_c^{(i)} w_{i_1}^+ w_{i_2}^+ \cdots w_{i_{n_+}}^+, \tag{39}$$

where $\widetilde{\mathbf{b}}^{(i)}$ is the vector

$$\widetilde{\mathbf{b}}^{(i)} = \mathbf{b}_0^{(i)} \ + \ (d_{i_1} - N_{i_1}^{(i)} k_i)\mathbf{b}_{i_1}^{(i)} \ + \sum_{\ell=1,\ \ell \neq i_1}^{k_i} d_\ell \mathbf{b}_\ell^{(i)}.$$

Example 2 of the Appendix clarifies the construction above.

4.5.2 L'_+ is commutatively equivalent to L_+

The following lemmas are intermediate steps to prove that L'_+ is commutatively equivalent to L_+. Their proofs are very similar to the one presented in [6]. For the sake of completeness, we report them in the Appendix.

Let $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be a simple set that appears in the decomposition of B_+ and let

$$L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}) = \varphi(B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}))$$

be the image under φ of the set $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$.

Moreover, let $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be the regular language associated with $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined by the Equation (39).

Lemma 8 *$L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ and $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ are commutatively equivalent.*

Lemma 9 *All the languages $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ are pairwise disjoint.*

Let L'_+ be the language over A:

$$L'_+ = \bigcup L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}), \tag{40}$$

given by the union of all the regular languages defined by the Equation (39).

Theorem 4 *The language L'_+ is regular and it is commutatively equivalent to L_+. Moreover every word of L'_+ has a suffix of length β that cannot be a factor of any word of $u_1^* \cdots u_k^*$.*

Proof. By definition the language L'_+ is regular. By Proposition 4, the languages $L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ gives a partition of L_+. By Lemma 9, the languages $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ give a partition of L'_+. By Lemma 8, every language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ is commutatively equivalent to $L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$. Then $L'_+ \sim L_+$ follows by Lemma 1.

By definition, every word of $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ ends with a code word of (38). Such code word has length β and cannot be factor of any word of $u_1^* \cdots u_k^*$. (see also Remark 2). □

4.6 The clusterization of $L_{a,c}$

In this section, we construct a regular language $L'_{a,c}$ which is commutatively equivalent to $L_{a,c}$. We use a technique very similar to that of Section 4.5.

4.6.1 The definition of $L'_{a,c}$

We recall that, according to point (3) of (24), $L_{a,c} = \varphi(B_{a,c})$, where $B_{a,c}$ is the union of all the simple sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ satisfying the following property:

3.1. there exist two indices ℓ_a and ℓ_c with $1 \leq \ell_a, \ell_c \leq k_i$ such that $\epsilon_{\ell_a} = \epsilon_{\ell_c} = +$ and $\varphi(N_{\ell_a}^{(i)} \mathbf{b}_{\ell_a}^{(i)}) \in a^+$, $\varphi(N_{\ell_a}^{(i)} \mathbf{b}_{\ell_c}^{(i)}) \in c^+$;

3.2. no index ℓ exists such that $\epsilon_\ell = +$ and $\varphi(N_\ell^{(i)} \mathbf{b}_\ell^{(i)})$ contains at least two distinct letters;

Let $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be a simple set that appears in the decomposition of $B_{a,c}$. For the sake of simplicity, denote it by $\mathcal{B}$. We want now to associate with $\mathcal{B}$ a regular language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$. For this purpose, it is useful to remind the definition of $\mathcal{B}$. By (11), (17), and (18), the simple set $\mathcal{B}$ has the form:

$$\mathcal{B} = \{\mathbf{b}_0^{(i)} \ + \ \sum_{\ell=1}^{k_i} d_\ell \mathbf{b}_\ell^{(i)} \ + \ \sum_{\ell=1}^{s} x_{i_\ell} N_{i_\ell}^{(i)} \mathbf{b}_{i_\ell}^{(i)} \ \mid \ x_{i_\ell} \geq 1\}, \tag{41}$$

where:

- the vectors $\mathbf{b}_0^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$ that appear in (41) form the representation of B_i and k_i is the dimension of B_i; moreover $s \geq 1$;
- there exist two indices i_{ℓ_a} and i_{ℓ_c} with $1 \leq i_{\ell_a}, i_{\ell_c} \leq k_i$ such that $d_{i_{\ell_a}} \geq N_{i_{\ell_a}}^{(i)}(k_i + 1)\beta$ and $d_{i_{\ell_c}} \geq N_{i_{\ell_c}}^{(i)}(k_i + 1)\beta$.

For the sake of simplicity, we denote i_{ℓ_a} and i_{ℓ_c} by i_1 and i_2, respectively.

According to the enumeration (14) of the generators of B_i defined in Section 4.1, we can display the generators $\mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$of B_i as:

$$\mathbf{b}_{a,1}^{(i)}, \ldots, \mathbf{b}_{a,n_a}^{(i)}, \tag{42}$$

$$\mathbf{b}_{c,1}^{(i)}, \ldots, \mathbf{b}_{c,n_c}^{(i)}, \tag{43}$$

$$\mathbf{b}_{+,1}^{(i)}, \ldots, \mathbf{b}_{+,n_+}^{(i)}, \tag{44}$$

where $n_a, n_c, n_+ \in \mathbb{N}$ and

- $\forall\ \ell = 1, \ldots, n_+,\ \varphi(\mathbf{b}_{+,\ell}^{(i)})$ contains at least two distinct letters;
- $\forall\ \ell = 1, \ldots, n_a,\ \varphi(\mathbf{b}_{a,\ell}^{(i)}) \in a^+$,
- $\forall\ \ell = 1, \ldots, n_c,\ \varphi(\mathbf{b}_{c,\ell}^{(i)}) \in c^+$.

Thus, by (42), (43), the formal sum of vectors $\sum_{\ell=1}^{s} x_{i_\ell} N_{i_\ell}^{(i)} \mathbf{b}_{i_\ell}^{(i)}$, that appears in the expression (41) can be rewritten as:

$$\sum_{\ell=1}^{p_a} x_{a,i_\ell} N_{a,i_\ell}^{(i)} \mathbf{b}_{a,i_\ell}^{(i)} \ + \ \sum_{\ell=1}^{p_c} y_{c,i_\ell} N_{c,i_\ell}^{(i)} \mathbf{b}_{c,i_\ell}^{(i)}, \tag{45}$$

where we have:

– $1 \leq p_a \leq n_a,\ 1 \leq p_c \leq n_c$,

– $N^{(i)}_{a,\ell}$ and $N^{(i)}_{c,\ell}$ denote the coefficients (17) of $\mathbf{b}^{(i)}_{a,\ell}$ and of $\mathbf{b}^{(i)}_{c,\ell}$, respectively, and x_{a,i_ℓ} and y_{c,i_ℓ} are free variables over $\mathbb{N}_+$.

This implies that the set $\mathcal{B}$ can be rewritten as:

$$\{\mathbf{b}^{(i)} + \sum_{\ell=1}^{p_a} x_{a,i_\ell} N_{a,i_\ell} \mathbf{b}^{(i)}_{a,i_\ell} + \sum_{\ell=1}^{p_c} y_{c,i_\ell} N_{c,i_\ell} \mathbf{b}^{(i)}_{c,i_\ell} \mid x_{a,i_\ell}, y_{c,i_\ell} \geq 1\}. \tag{46}$$

with $\mathbf{b}^{(i)} = \mathbf{b}^{(i)}_0 + \sum_{\ell=1}^{k_i} d_\ell \mathbf{b}^{(i)}_\ell$. By using the same argument of the proof of Lemma 4, we can construct a word w of the following type.

Property 1 *The word w fulfills the following properties:*
– $\psi(a^\beta c^\beta) = \psi(w)$,
– $w = uv$, where u and v are words of length β, with $u \neq v$, both containing 2 distinct letters,
– u cannot be factor of any word of $u_1^* \cdots u_k^*$.

Now let us define the language $\mathrm{L}^{(i)}_a$ as:

$$\mathrm{L}^{(i)}_a = wL_{a,1}wL_{a,2}w \cdots wL_{a,n_a}, \tag{47}$$

where:

- n_a is the length of the list (42);
- w is the word defined above and it occurs n_a times in $\mathrm{L}^{(i)}_a$;
- For every $j = 1, \ldots, n_a$, we set $L_{a,j} = (a^\beta)^+$ if the vector $\mathbf{b}^{(i)}_{a,j}$ of the list (42) appears in the sum (45); otherwise it is equal to 1_{A^*}.

Similarly, let us define the language $\mathrm{L}^{(i)}_c$ as:

$$\mathrm{L}^{(i)}_c = wL_{c,1}wL_{c,2}w \cdots wL_{c,n_c}, \tag{48}$$

where:

- n_c is the length of the list (43);
- w is the word defined above and it occurs n_c times in $\mathrm{L}^{(i)}_c$;

- For every $j = 1, \ldots, n_c$, we set $L_{c,j} = (c^\beta)^+$ if the vector $\mathbf{b}_{c,j}^{(i)}$ of the list (43) appears in the sum (45); otherwise it is equal to 1_{A^*}.

Remark 6 Every vector $N_{i_\ell}^{(i)} \mathbf{b}_{a,i_\ell}^{(i)}$ of $\mathcal{B}$ is univocally represented by the presence into the language $\mathrm{L}_a^{(i)}$ of the factor $L_{a,i_\ell} = (a^\beta)^+$ in the corresponding *slot* i_ℓ. The same remark holds for $\mathrm{L}_c^{(i)}$.

Finally we associate with $\mathcal{B}$ the regular language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined as:

$$\varphi(\widetilde{\mathbf{b}}^{(i)}) w^{(k_i - n_a - n_c)} \mathrm{L}_a^{(i)} \mathrm{L}_c^{(i)} w, \tag{49}$$

where $\widetilde{\mathbf{b}}$ is the vector

$$\widetilde{\mathbf{b}}^{(i)} = \mathbf{b}_0^{(i)} + (d_{i_1} - N_{i_1}^{(i)}(k_i + 1))\mathbf{b}_{i_1}^{(i)} + (d_{i_2} - N_{i_2}^{(i)}(k_i + 1))\mathbf{b}_{i_2}^{(i)} + \sum_{\ell=1,\ \ell \neq i_1, i_2}^{k_i} d_\ell \mathbf{b}_\ell^{(i)}.$$

Example 3 of the Appendix clarifies the construction above.

4.6.2 $L'_{a,c}$ is commutatively equivalent to $L_{a,c}$

The following lemmata are intermediate steps to prove that $L'_{a,c}$ is commutatively equivalent to $L_{a,c}$. Their proofs are very similar to those of Lemma 8 and Lemma 9 respectively and therefore they are omitted.

Let $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be a simple set of the decomposition of $B_{a,c}$ and let $L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}) = \varphi(B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}))$ be the image under φ of $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$.

Moreover, let $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ be the regular language associated with $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined by (49).

Lemma 10 *$L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ and $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ are commutatively equivalent.*

Lemma 11 *All the languages $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ are pairwise disjoint.*

Let $L'_{a,c}$ be the language over A:

$$L' = \bigcup L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i}), \tag{50}$$

given by the union of all the regular languages defined by the Equation (49). The proof of the following theorem is based upon Lemma 10 and Lemma 11 and follows the very same scheme of that of Theorem 4; thus it is omitted.

Theorem 5 *The language $L'_{a,c}$ is regular and it is commutatively equivalent to $L_{a,c}$. Moreover every word of $L'_{a,c}$ has a suffix of length 2β that cannot be a factor of any word of $u_1^* \cdots u_k^*$.*

4.7 The clusterization of L_a and of L_c

Let $L_a = \varphi(B_a)$ and $L_c = \varphi(B_c)$ be the languages defined in Section 4.4. The aim of this section is to prove the following theorems.

Theorem 6 *There exists a regular language L'_a which is commutatively equivalent to L_a. Every word of L'_a ends with a word u of length 2β, where β is the constant defined in Section 4.2, and u is a factor of some word of $u_1^* \cdots u_k^*$.*

Theorem 7 *There exists a regular language L'_c which is commutatively equivalent to L_c. Every word of L'_c ends with a word u of length 2β, where β is the constant defined in Section 4.2, and u is a factor of some word of $u_1^* \cdots u_k^*$.*

We present only the proof of Theorem 6, since the proof of Theorem 7 is exactly the same. As a general remark, we point out to the reader that the proof of Theorem 6 will be essentially based upon the main result of [7].

Now, it is useful to remind the definition of L_a as well as some of its properties. According to point (4) of (24), the semi-simple set B_a has been defined as the union of the pairwise disjoint simple sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ of dimension ≥ 1

$$\{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m : x_1, \ldots, x_m \geq 2\beta\}, \tag{51}$$

where $\mathbf{b}_0, \ldots, \mathbf{b}_m$ are vectors of $\mathbb{N}^k$ such that $\mathbf{b}_1, \ldots, \mathbf{b}_m$ are linearly independent and, for every $\ell = 1, \ldots, m$, $\varphi(\mathbf{b}_\ell) \in a^+$. For the sake of simplicity we still call them (with a minor abuse of terminology) the vectors of the representation of $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$. It is also useful to remark that, by (18), the coefficients $x_1, \ldots, x_m$ in (51), are larger than 2β.

In particular, by (27), one has:

$$\forall\ u \in L_a, \quad |u|_c < \mathrm{Max}_c, \tag{52}$$

where Max_c is defined in (26). For the sake of simplicity, we fix once for all an enumeration of the simple sets $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ of the decomposition of B_a. Thus we can denote an arbitrary set of such family as B^a_ℓ and B_a can be written as:

$$B_a = B^a_1 \cup \cdots \cup B^a_s, \quad s \geq 1. \tag{53}$$

The following fact is easily proved.

Lemma 12 *Let B and $\bar{B}$ be two simple sets of the decomposition (53), and let $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ and $\bar{\mathbf{b}}_0, \bar{\mathbf{b}}_1, \ldots, \bar{\mathbf{b}}_{m'}$, be the representations of B and $\bar{B}$ respectively. Assume that $|\varphi(\mathbf{b}_0)|_c \neq |\varphi(\bar{\mathbf{b}}_0)|_c$. Let L and $\bar{L}$ be two languages commutatively equivalent to $\varphi(B)$ and to $\varphi(\bar{B})$, respectively. Then L and $\bar{L}$ are disjoint.*

Proof. By (51), for every $\mathbf{b} \in B$, $|\varphi(\mathbf{b})|_c = |\varphi(\mathbf{b}_0)|_c$ and, for every $\bar{\mathbf{b}} \in \bar{B}$, $|\varphi(\bar{\mathbf{b}})|_c = |\varphi(\bar{\mathbf{b}}_0)|_c$. This implies $\varphi(B) \cap \varphi(\bar{B}) = \emptyset$. Then the claim now follows from $L \sim \varphi(B)$ and $\bar{L} \sim \varphi(\bar{B})$. □

Remark 7 The reader can easily see that, in the general case of an alphabet with an arbitrary number of letters, the general version of Lemma 12 consists in separating the languages not only w.r.t. the number of occurrences of the letter c but also w.r.t. the numbers of occurrences of all the letters different from a as well as their relative positions in the word $\varphi(\mathbf{b}_0)$.

Let us consider, for every $\ell = 1, \ldots, s$, the simple set B^a_ℓ of (53) and denote by t_ℓ the number of occurrences of the letter c in the word $\varphi(\mathbf{b}^{(a)}_0)$, where $\mathbf{b}^{(a)}_0$ is the constant vector of the representation of B^a_ℓ. From (51) and (52), it follows that, for every $u \in \varphi(B^a_\ell)$, $|u|_c = t_\ell < \mathrm{Max}_c$. For a given $t = 0, \ldots, \mathrm{Max}_c - 1$, let $B^a_{t,1}, \ldots, B^a_{t,s_t}$ be the subsequence of all simple sets B^a_ℓ in the decomposition (53) such that $t_\ell = t$. Define the semi-simple set C_t as:

$$C_t = B^a_{t,1} \cup \cdots \cup B^a_{t,s_t}, \tag{54}$$

and let $L^a_t = \varphi(C_t)$.

Lemma 13 *Assume that, for every $t = 0, \ldots, Max_c - 1$, there exists a regular language L'^a_t which is commutatively equivalent to L^a_t. Then there exists a regular language L'_a which is commutatively equivalent to L_a.*

Proof. By (53) and (54), one has $B_a = \bigcup_{t=0}^{Max_c - 1} C_t$ so that $L_a = \varphi(B_a) = \bigcup_{t=0}^{Max_c - 1} L^a_t$. Moreover, by Proposition 2, the sets L^a_t are pairwise disjoint. Let us consider the regular language $L'_a = \bigcup_{t=0}^{Max_c - 1} L'^a_t$. Observe now that the languages L'^a_t, with $0 \leq t < \mathrm{Max}_c$, are pairwise disjoint. Indeed, this follows by applying Lemma 12, and taking into account that, for every $t = 0, \ldots, \mathrm{Max}_c - 1$, $L'^a_t \sim L^a_t$. The claim follows by applying Lemma 1. □

For Lemma 13, to prove Theorem 6, it is enough to show the following.

Theorem 8 *Let t be a fixed integer with $0 \leq t \leq Max_c - 1$. There exists a regular language L'^a_t such that the following two conditions hold:*

(i) L'^a_t is commutatively equivalent to L^a_t;

(ii) Every word of L'^a_t ends with a word of length 2β, where β is the constant defined in Section 4.2, which is a factor of some word of $u_1^* \cdots u_k^*$.

In the sequel, we will assume that $t \geq 1$. Indeed, if $t = 0$, the language L^a_t is obviously regular and nothing has to be proved.

In order to prove Theorem 8, let us show some preliminary results on the structure of the language L^a_t. By (54), one has that $L^a_t = \bigcup_{j=1}^{s_t} L^a_{t,j}$, where

$$\forall\, j = 1, \ldots, s_t, \quad L^a_{t,j} = \varphi(B^a_{t,j}). \tag{55}$$

The following lemma holds.

Lemma 14 *Let $L^a_{t,j}$, with $j = 1, \ldots, s_t$, be a language of (55). There exist a word v_j and a non-negative integer m_j such that:*

$$L^a_{t,j} \quad \subseteq \quad \underbrace{a^*ca^*ca^*c \cdots ca^*}_{m_j-times} v_j, \tag{56}$$

where

– v_j is a word of length 2β, where β is the constant defined in Section 4.2, and it is a factor of some word of $u_1^* \cdots u_k^*$;

– the number m_j of occurrences of the symbol c in the right-side bounded expression of (56) is such that $t = m_j + |v_j|_c$.

Proof. Let $B^a_{t,j}$ be the simple set such that $L^a_{t,j} = \varphi(B^a_{t,j})$ and let $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ be the representation of $B^a_{t,j}$. For every $i = 1, \ldots, m$, let $\mathbf{b}_i = (b_{i1}, \ldots, b_{ik})$ and let r_i be the largest index ℓ, with $1 \leq \ell \leq k$, such that $b_{ik} > 0$. Let $R = \max_{1 \leq i \leq m} r_i$. Then, for every $\mathbf{b} \in B^a_{t,j}$, we get

$\mathbf{b} = (b_{01} + \sum_{h=1}^m x_h b_{h1}, \ldots, b_{0R} + \sum_{h=1}^m x_h b_{hR}, b_{0R+1}, b_{0R+2}, \ldots, b_{0k})$,

so that

$\varphi(\mathbf{b}) = u_1^{b_{01}+\sum_{h=1}^m x_h b_{h1}} \cdots u_R^{b_{0R}+\sum_{h=1}^m x_h b_{hR}} u_{R+1}^{b_{0R+1}} u_{R+2}^{b_{0R+2}} \cdots u_k^{b_{0k}}$.

Since $B^a_{t,j}$ has the form (51), we have $x_R \geq 2\beta$ and $u_R \in a^+$, which implies that the word $\mathrm{v}_j = a^{2\beta} u_{R+1}^{b_{0R+1}} u_{R+2}^{b_{0R+2}} \cdots u_k^{b_{0k}}$ is a suffix of $\varphi(\mathbf{b})$. The proof is complete. □

Let $\mathcal{U}$ be the set obtained by collecting all the words v_j, with $j = 1, \ldots, s_t$. Let $u \in \mathcal{U}$ and denote by m_u the corresponding number of occurrences of c that appears in the rightside expression of (56). Let

$$L^a_{t,u,1}, \ldots, L^a_{t,u,s_u}, \tag{57}$$

be the subsequence of all the languages $L^a_{t,j}$, with $j = 1, \ldots, s_t$ such that the corresponding word v_j is u. Moreover, set $L^a_{t,u} = L^a_{t,u,1} \cup \cdots \cup L^a_{t,u,s_u}$.

Lemma 15 *Assume that, for every $u \in \mathcal{U}$, there exists a regular language $L'^a_{t,u}$ such that $L'^a_{t,u}$ is commutatively equivalent to $L^a_{t,u}$ and*

$$L'^a_{t,u} \subseteq \underbrace{a^*ca^*ca^*c \cdots ca^*}_{m_u-times} u.$$

Then there exists a regular language L'^a_t which is commutatively equivalent to L^a_t. Moreover, every word of L'^a_t ends with a word of length 2β, which is a factor of some word of $u_1^* \cdots u_k^*$.

Proof. Let us consider the regular language $L'^a_t = \bigcup_{u \in \mathcal{U}} L'^a_{t,u}$. Observe now that, since all the words of $\mathcal{U}$ have the same length 2β, if u_1, u_2 are distinct words of $\mathcal{U}$, $L'^a_{t,u_1} \cap L'^a_{t,u_2} = \emptyset$. Since L^a_t is partitioned as $L^a_t = \bigcup_{u \in \mathcal{U}} L^a_{t,u}$, $L'^a_t \sim L^a_t$ follows from the latter and the hypothesis, by applying Lemma 1. Moreover, by construction, every word of L'^a_t ends with some word u of $\mathcal{U}$, which is a factor of some word of $u_1^* \cdots u_k^*$. □

By Lemma 15, it is enough to prove Theorem 8 with respect to the family of languages (57). Thus, let us fix, once for all, a word u in $\mathcal{U}$ and let

$$B^a_{t,u,1}, \ldots, B^a_{t,u,s_u}, \tag{58}$$

be the subsequence of the simple sets $B^a_{t,u,\ell}$ of (54) such that, for every language $L^a_{t,u,\ell}$ of (57), $\varphi(B^a_{t,u,\ell}) = L^a_{t,u,\ell}$. Let us consider the map

$$\widehat{\varphi} : \mathbb{N}^{t+1} \longrightarrow \underbrace{a^*ca^*ca^*ca^* \cdots a^*ca^*}_{t-times},$$

from $\mathbb{N}^{t+1}$ into the language $a^*ca^*ca^*ca^* \cdots a^*ca^*$, where the symbol c occurs t times in the bounded expression above, defined as: for every $\mathbf{v} = (v_1, \ldots, v_{t+1}) \in \mathbb{N}^{t+1}$ $\widehat{\varphi}(\mathbf{v}) = \widehat{\varphi}(v_1, \ldots, v_{t+1}) = a^{v_1}ca^{v_2}c \cdots ca^{v_{t+1}}$. By definition, the map $\widehat{\varphi}$ is injective on its domain $\mathbb{N}^{t+1}$. Given a vector $\mathbf{v}$ of $\mathbb{N}^{t+1}$, the weight of $\mathbf{v}$ is the number $|\mathbf{v}| = v_1 + \cdots + v_{t+1}$. Given two vectors $\mathbf{b}$ and $\mathbf{b}'$ of $\mathbb{N}^{t+1}$, then one immediately checks

$$|\mathbf{b}| = |\mathbf{b}'| \ \Rightarrow \ \psi(\widehat{\varphi}(\mathbf{b})) = \psi(\widehat{\varphi}(\mathbf{b}')). \tag{59}$$

The proof of the following result is postponed in the Appendix.

Proposition 5 *For every simple set $B^a_{t,u,\ell}$, with $1 \leq \ell \leq s_u$, there exists a simple set $\widehat{B}^a_{t,u,\ell}$ of $\mathbb{N}^{t+1}$ such that $\widehat{\varphi}(\widehat{B}^a_{t,u,\ell}) = \varphi(B^a_{t,u,\ell})$ and $\widehat{\varphi}$ is injective on $\widehat{B}^a_{t,u,\ell}$. Moreover, there exists a vector $\mathbf{b}_u \in \mathbb{N}^{t+1}$ such that, for every $\ell = 1, \ldots, s_u$, the set $\widehat{B}^a_{t,u,\ell}$ has the form*

$$\widehat{B}^a_{t,u,\ell} = \widehat{D}^a_{t,u,\ell} \times 0^{t-m_u} + \mathbf{b}_u,$$

where $\widehat{D}^a_{t,u,\ell}$ is a simple set of $\mathbb{N}^{m_u+1}$.

Proposition 6 *There exists a semi-simple set $\widehat{C}^a_{t,u}$ of $\mathbb{N}^{t+1}$ where $\widehat{\varphi}(\widehat{C}^a_{t,u}) = L^a_{t,u}$ and $\widehat{\varphi}$ is injective on $\widehat{C}^a_{t,u}$. Moreover, there exists a vector $\mathbf{b}_u \in \mathbb{N}^{t+1}$ such that $\widehat{C}^a_{t,u}$ has the form $\widehat{C}^a_{t,u} = \widehat{E}^a_{t,u} \times 0^{t-m_u} + \mathbf{b}_u$, where $\widehat{E}^a_{t,u}$ is a semi-simple set of $\mathbb{N}^{m_u+1}$.*

Proof. By applying Proposition 5, for every $\ell = 1, \ldots, s_t$, there exists a simple set $\widehat{B}^a_{t,u,\ell}$ of $\mathbb{N}^{t+1}$ such that $\widehat{\varphi}(\widehat{B}^a_{t,u,\ell}) = \varphi(B^a_{t,u,\ell})$. Let $\widehat{C}^a_{t,u} = \bigcup_{\ell=1}^{s_t} \widehat{B}^a_{t,u,\ell}$. Observe now that the sets $\widehat{B}^a_{t,u,\ell}$ for $1 \leq \ell \leq s_t$ are pairwise disjoint. This immediately comes from the latter and the fact that the languages $\varphi(B^a_{t,u,\ell})$, with $1 \leq \ell \leq s_t$, are pairwise disjoint. Finally, one checks that $\widehat{\varphi}$ is injective on $\widehat{C}^a_{t,u}$. Indeed, let $\mathbf{b}_1, \mathbf{b}_2 \in \widehat{C}^a_{t,u}$, with $\widehat{\varphi}(\mathbf{b}_1) = \widehat{\varphi}(\mathbf{b}_2)$. If $\mathbf{b}_1, \mathbf{b}_2 \in B^a_{t,u,\ell}$, with $1 \leq \ell \leq s_t$, then the claim follows from Proposition 5. If $\mathbf{b}_1 \in B^a_{t,u,\ell}$, and $\mathbf{b}_2 \in B^a_{t,u,\ell'}$, with $\ell \neq \ell'$, then the claim follows from the fact that the languages $\varphi(B^a_{t,u,\ell})$, with $1 \leq \ell \leq s_t$, are pairwise disjoint. The second part of the claim follows from the second part of the claim of Proposition 5. □

By Proposition 6, there exists a semi-simple set $\widehat{C}^a_{t,u}$ of $\mathbb{N}^{t+1}$ such that $\widehat{\varphi}(\widehat{C}^a_{t,u}) = L^a_{t,u}$ and $\widehat{C}^a_{t,u}$ has the form $\widehat{C}^a_{t,u} = \widehat{E}^a_{t,u} \times 0^{t-m_u} + \mathbf{b}_u$, where $\widehat{E}^a_{t,u}$ is a semi-simple set of $\mathbb{N}^{m_u+1}$.

By Theorem 1 of [7] applied to the semi-simple set $\widehat{E}^a_{t,u}$ of $\mathbb{N}^{m_u+1}$, there exists a recognizable semi-simple set $\widehat{E}'^a_{t,u}$ of $\mathbb{N}^{m_u+1}$, which is commutatively equivalent to $\widehat{E}^a_{t,u}$. This is equivalent to say that:

1) $\widehat{E}'^{a}_{t,u}$ is a finite union of pairwise disjoint simple sets of $\mathbb{N}^{m_u+1}$ of the form $\mathbf{v}_0 + \{\mathbf{v}_1, \dots, \mathbf{v}_r\}^{\oplus}$, where $r \geq 0$[1] and, for every $\ell = 1, \dots, r$, exactly one component of $\mathbf{v}_\ell$ is not null and different vectors $\mathbf{v}_\ell$, with $1 \leq \ell \leq r$, have a different non-null component;

2) there exists a bijection $f : \widehat{E}^{a}_{t,u} \longrightarrow \widehat{E}'^{a}_{t,u}$ between $\widehat{E}^{a}_{t,u}$ and $\widehat{E}'^{a}_{t,u}$ such that, for every $\mathbf{v} \in \widehat{E}^{a}_{t,u}$, $|f(\mathbf{v})| = |\mathbf{v}|$.

Let us now consider the subset of $\mathbb{N}^{t+1}$ $\widehat{C}'^{a}_{t,u} = \widehat{E}'^{a}_{t,u} \times 0^{t-m_u} + \mathbf{b}_u$. One easily checks that $\widehat{C}'^{a}_{t,u}$ is a recognizable set of $\mathbb{N}^{t+1}$ and $\widehat{C}'^{a}_{t,u}$ is commutatively equivalent to $\widehat{C}^{a}_{t,u}$. Let us set $L'^{a}_{t,u} = \widehat{\varphi}(\widehat{C}'^{a}_{t,u})$. The following result holds.

Lemma 16 *$L'^{a}_{t,u}$ is a regular language and every word of $L'^{a}_{t,u}$ ends with the word u.*

Proof. We have $L'^{a}_{t,u} = \widehat{\varphi}(\widehat{C}'^{a}_{t,u})$, where $\widehat{C}'^{a}_{t,u} = \widehat{E}'^{a}_{t,u} \times 0^{t-m_u} + \mathbf{b}_u$. Since $\widehat{E}'^{a}_{t,u}$ is a semi-simple recognizable set of $\mathbb{N}^{m_u+1}$, $\widehat{C}'^{a}_{t,u}$ is a finite union of pairwise disjoint simple sets of the form $B' \times 0^{t-m_u} + \mathbf{b}_u$ where B' is a simple recognizable subset of $\mathbb{N}^{m_u+1}$. Therefore it is enough to prove the claim for every one of such set $B' \times 0^{t-m_u} + \mathbf{b}_u$.

Let $B' = \mathbf{v}_0 + \{\mathbf{v}_1, \dots, \mathbf{v}_r\}^{\oplus}$, where $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_r$ are the vectors of the representation of B'. Assume first that $r \geq 1$. Denote v_{ij} the j-th-component of the vector $\mathbf{v}_i$, with $0 \leq i \leq r$ and $1 \leq j \leq m_u + 1$. Remind that, for every $i = 1, \dots, r$, exactly one component of $\mathbf{v}_i$ is not null. We can assume that the unique non null component of $\mathbf{v}_i$ is the i-th component v_{ii}, with $i = 1, \dots, r$, the other cases being completely similar. By the latter, the arbitrary vector of $B' \times 0^{t-m_u} + \mathbf{b}_u$ is written as

$$\mathbf{v} = (v_{01} + x_1 v_{11}, \dots, v_{0r} + x_r v_{rr}, \underbrace{0, 0, \dots, 0}_{m_u+1-r}) + \mathbf{b}_u, \quad x_1, \dots, x_r \geq 0.$$

Since, by Eq. (71), one has $\mathbf{b}_u = (\underbrace{0, 0, \dots, 0}_{m_u}, \alpha_1, \dots, \alpha_{n+1})$, with $n = |u|_c$, and the numbers α_ℓ are such that $u = a^{\alpha_1} c a^{\alpha_2} c \cdots c a^{\alpha_n} c a^{\alpha_{n+1}}$, the image under the map $\widehat{\varphi}$ of $\mathbf{v}$ is $\widehat{\varphi}(\mathbf{v}) = a^{v_{01}+x_1 v_{11}} c \cdots c a^{v_{0r}+x_r v_{rr}} \underbrace{cc \cdots c}_{m_u+1-r} u$.

Hence the image under the map $\widehat{\varphi}$ of the set $B' \times 0^{t-m_u} + \mathbf{b}_u$ is the regular language $a^{v_{01}}(a^{v_{11}})^* c \cdots c a^{v_{0r}}(a^{v_{rr}})^* \underbrace{cc \cdots c}_{m_u+1-r} u$.

[1] it is understood that if $r = 0$, the corresponding set of generators is empty

The case $r = 0$ is similarly treated. This completes the proof. $\square$

Now we show that the regular language $L'^{a}_{t,u} = \widehat{\varphi}(\widehat{C}'^{a}_{t,u})$ allows us to complete the proof of Theorem 8.

Proof of Theorem 8: Let $L'^{a}_{t,u} = \widehat{\varphi}(\widehat{C}'^{a}_{t,u})$ be the language defined above. By Lemma 16, $L'^{a}_{t,u}$ is regular and every word of $L'^{a}_{t,u}$ ends with the word u. Thus, in order to complete the proof, we need to show that $L'^{a}_{t,u}$ is commutatively equivalent to $L^{a}_{t,u}$. Since $\widehat{C}'^{a}_{t,u}$ and $\widehat{C}^{a}_{t,u}$ are commutatively equivalent, there exists a bijection $f : \widehat{C}^{a}_{t,u} \longrightarrow \widehat{C}'^{a}_{t,u}$ from $\widehat{C}^{a}_{t,u}$ onto $\widehat{C}'^{a}_{t,u}$ such that, for every $\widehat{\mathbf{v}} \in \widehat{C}^{a}_{t,u}$, $|f(\widehat{\mathbf{v}})| = |\widehat{\mathbf{v}}|$. Let $v \in L^{a}_{t,u}$. By Proposition 6, there exists exactly one vector $\widehat{\mathbf{b}} \in \widehat{C}^{a}_{t,u}$ with $v = \widehat{\varphi}(\widehat{\mathbf{b}})$. If we consider the image of $\widehat{\mathbf{b}}$ under the map f we have $|\widehat{\mathbf{b}}| = |f(\widehat{\mathbf{b}})|$, and, by (59), one gets

$$\psi(\widehat{\varphi}(\widehat{\mathbf{b}})) = \psi(v) = \psi(\widehat{\varphi}(f(\widehat{\mathbf{b}}))). \tag{60}$$

By Proposition 6, $\widehat{\varphi}^{-1} : L^{a}_{t,u} \longrightarrow \widehat{C}^{a}_{t,u}$ is a bijection from $L^{a}_{t,u}$ onto $\widehat{C}^{a}_{t,u}$. Taking into account that f is a bijection from $\widehat{C}^{a}_{t,u}$ onto $\widehat{C}'^{a}_{t,u}$, and $\widehat{\varphi} : \widehat{C}'^{a}_{t,u} \longrightarrow L'^{a}_{t,u}$ is a bijection from $\widehat{C}'^{a}_{t,u}$ onto $L'^{a}_{t,u}$, one has that the map $\widehat{\varphi}^{-1} f \widehat{\varphi} : L^{a}_{t,u} \longrightarrow L'^{a}_{t,u}$ is a bijection from $L^{a}_{t,u}$ onto $L'^{a}_{t,u}$. Moreover, by (60), such maps preserve the Parikh vectors. This concludes the proof. $\square$

4.8 Proof of Theorem 1

Let us consider the regular languages $L'_{+}, L'_{a,c}, L'_{a}$, and L'_{c} defined by Theorem 4, Theorem 5, Theorem 6, and Theorem 7, respectively. The following lemma is instrumental for the proof of Theorem 1. Its proof is a technical consequence of the results proved in the previous sections and is postponed in the Appendix.

Lemma 17 *The languages L'_{+}, $L'_{a,c}$, L'_{a}, and L'_{c} are pairwise disjoint.*

We are now able to prove our main result.

Proof of Theorem 1: Let $L = \varphi(B)$ be a bounded semi-linear language contained in $u_1^* \cdots u_k^*$ and described, via the map φ, by a semi-simple set B. If B is finite the claim is trivial. Assume that B is not finite and apply to B the algorithm described in the previous sections. By Corollary 1, one gets the decomposition (25) $L = L_- \cup L_+ \cup L_{a,c} \cup L_a \cup L_c$ of L into semilinear languages. By Theorem 4, Theorem 5, Theorem 6, and Theorem 7, respectively, there exist regular languages $L'_+, L'_{a,c}, L'_a$, and L'_c such that $L'_+ \sim L_+$, $L'_{a,c} \sim L_{a,c}$, $L'_a \sim L_a$, and $L'_c \sim L_c$. By Lemma 17, the languages $L'_+, L'_{a,c}, L'_a$, and L'_c, are pairwise disjoint, so that, by Lemma 1, we have $L_+ \cup L_{a,c} \cup L_a \cup L_c \ \sim \ L'_+ \cup L'_{a,c} \cup L'_a \cup L'_c$. By Lemma 2, there exists a finite set of words L'_- such that the languages $L = L_- \cup L_+ \cup L_{a,c} \cup L_a \cup L_c$ and $L'_- \cup L'_+ \cup L'_{a,c} \cup L'_a \cup L'_c$ are commutatively equivalent. Finally, we observe that every step of the construction of L' is effective. □

Acknowledgements. We thank the anonymous referee for her (his) help in improving a previous version of this paper. In particular, we thank her (him) for pointing out Lemma 2 that simplifies some parts of the proof.

References

[1] J. Berstel, D. Perrin, C. Reutenauer, *Codes and Automata*, Encyclopedia of Mathematics and its Applications No. 129, Cambridge University Press, Cambridge, (2009).

[2] F. D'Alessandro, B. Intrigila, S. Varricchio, On the structure of the counting function of context-free languages, Theoret. Comput. Sci. **356**, 104–117 (2006).

[3] F. D'Alessandro, B. Intrigila, S. Varricchio, The Parikh counting functions of sparse context-free languages are quasi-polynomials, Theoret. Comput. Sci. **410**, 5158–5181 (2009).

[4] F. D'Alessandro, B. Intrigila, The commutative equivalence of bounded context-free and regular languages, in International Conference on words and formal languages, Words 2011, Electronic Proceedings in Theoretical Computer Science, p. 1-21, doi: 10.4204/EPTCS.63 (2011).

[5] F. D'Alessandro, B. Intrigila, S. Varricchio, Quasi-polynomials, Semilinear set, and Linear Diophantine equations, Theoret. Comput. Sci. **416**, 1–16 (2012).

[6] F. D'Alessandro, B. Intrigila, On the commutative equivalence of bounded context-free and regular languages: the code case, Theoret. Comput. Sci. **562**, 304–319 (2015).

[7] F. D'Alessandro, B. Intrigila, On the commutative equivalence of semilinear sets of $\mathbb{N}^k$, Theoret. Comput. Sci. **562**, 476–495 (2015).

[8] S. Eilenberg, M. -P. Schützenberger, Rational sets in commutative monoids, J. of Algebra **13**, 173–191 (1969).

[9] P. Flajolet, Analytic models and ambiguity of context-free languages, Theoret. Comput. Sci. **49**, 283–309 (1987).

[10] S. Ginsburg, *The mathematical theory of context-free languages*, Mc Graw- Hill, New York, (1966).

[11] J. Honkala, Decision problems concerning thinness and slenderness of formal languages, Acta Inf. **35**, 625–636 (1998).

[12] Perrin D., *private communication,* (2014).

Appendix

Proof of lemma 8: Denote $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ by $\mathcal{B}$. Moreover denote the language $L(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ by $\mathcal{L}$ and the regular language $L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ by $\mathcal{L}'$. Let $u \in \mathcal{L}$. By Proposition 2 there exists exactly one vector $\mathbf{v} \in B$ such that $u = \varphi(\mathbf{v})$. Obviously one has $\mathbf{v} \in \mathcal{B}$. Hence, by (37), there exists exactly one tuple of positive integers $x_{i_1}, \ldots, x_{i_{p_a}}$, $y_{i_1}, \ldots, y_{i_{p_c}}$, $z_{i_1}, \ldots, z_{i_{p_+}}$, such that

$$\mathbf{v} = \mathbf{b} + \sum_{\ell=1}^{p_a} x_{i_\ell} N_{a,i_\ell} \mathbf{b}^{(i)}_{a,i_\ell} + \sum_{\ell=1}^{p_c} y_{i_\ell} N_{c,i_\ell} \mathbf{b}^{(i)}_{c,i_\ell} + \sum_{\ell=1}^{p_+} z_{i_\ell} N_{+,i_\ell} \mathbf{b}^{(i)}_{+,i_\ell}, \qquad (61)$$

with $\mathbf{b} = \mathbf{b}^{(i)}_0 + \sum_{\ell=1}^{k_i} d_\ell \mathbf{b}^{(i)}_\ell$. Let us consider the map $f : \mathcal{L} \longrightarrow \mathcal{L}'$ such that, for every $u \in \mathcal{L}$,

$$f(u) = \varphi(\widetilde{\mathbf{b}}) w_{i_1}^{k_i - n_a - n_c} \mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_1} w_{i_2}^{z_2} \cdots w_{i_{p_+}}^{z_{p_+}},$$

where

- $\widetilde{\mathbf{b}} = \mathbf{b}^{(i)}_0 + (d_{i_1} - N^{(i)}_{i_1} k_i) \mathbf{b}^{(i)}_{i_1} + \sum_{\ell=1,\ \ell \neq i_1}^{k_i} d_\ell \mathbf{b}^{(i)}_\ell$.
- $\mathrm{u}_a = w_{i_1} b_1 w_{i_1} b_2 \cdots w_{i_1} b_{n_a}$ is the word of $\mathrm{L}^{(i)}_a$ defined by the sequence $x_{i_1}, \ldots, x_{i_{p_a}}$, in the proper way: for every $j = 1, \ldots, n_a$:

$$b_j = \begin{cases} a^{\beta x_j} & \text{if } j = i_\ell, \\ 1_{A^*} & \text{otherwise.} \end{cases}$$

- $\mathrm{u}_c = w_{i_1} b_1 w_{i_1} b_2 \cdots w_{i_1} b_{n_c}$ is the word of $\mathrm{L}_c^{(i)}$ defined by the sequence $y_{i_1}, \ldots, y_{i_{p_c}}$ in the proper way: for every $j = 1, \ldots, n_c$:

$$b_j = \begin{cases} c^{\beta y_j} & \text{if } j = i_\ell, \\ 1_{A^*} & \text{otherwise.} \end{cases}$$

It is easily checked that f is well defined as a map from $\mathcal{L}$ to $\mathcal{L}'$. Our main task is to prove that f is a bijection from $\mathcal{L}$ to $\mathcal{L}'$ that preserves the Parikh vectors of words of $\mathcal{L}$. Let us prove that f is a bijection from $\mathcal{L}$ to $\mathcal{L}'$. From the definition of f, it is easily checked that f is a surjective map. Let us prove that f is injective. Consider another word $u' \in \mathcal{L}$ and assume $f(u) = f(u')$. As before, there exists exactly one tuple of positive integers $x'_{i_1}, \ldots, x'_{i_{p_a}}$, $y'_{i_1}, \ldots, y'_{i_{p_c}}$, $z'_{i_1}, \ldots, z'_{i_{p_+}}$, such that $u' = \varphi(\mathbf{v}')$ where:

$$\mathbf{v}' = \mathbf{b} + \sum_{\ell=1}^{p_a} x'_{i_\ell} N_{a,i_\ell} \mathbf{b}^{(i)}_{a,i_\ell} + \sum_{\ell=1}^{p_c} y'_{i_\ell} N_{c,i_\ell} \mathbf{b}^{(i)}_{c,i_\ell} + \sum_{\ell=1}^{p_+} z'_{i_\ell} N_{+,i_\ell} \mathbf{b}^{(i)}_{+,i_\ell}.$$

Then $f(u') = \varphi(\widetilde{\mathbf{b}}) w_{i_1}^{k_i - n_a - n_c} \mathrm{u}'_a \mathrm{u}'_c w_{i_1}^{z'_1} w_{i_2}^{z'_2} \cdots w_{i_{p_+}}^{z'_{p_+}}$, where u'_a and u'_c are determined by the sequences $x'_{i_1}, \ldots, x'_{i_{p_a}}$ and $y'_{i_1}, \ldots, y'_{i_{p_c}}$, respectively, similarly to u_a and u_c. From the equality $f(u) = f(u')$, one has

$$\mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_1} w_{i_2}^{z_2} \cdots w_{i_{p_+}}^{z_{p_+}} = \mathrm{u}'_a \mathrm{u}'_c w_{i_1}^{z'_1} w_{i_2}^{z'_2} \cdots w_{i_{p_+}}^{z'_{p_+}},$$

which immediately implies

$$x_{i_1} = x'_{i_1}, \ldots, x_{i_{p_a}} = x'_{i_{p_a}},\ y_{i_1} = y'_{i_1}, \ldots, y_{i_{p_c}} = y'_{i_{p_c}},\ z_{i_1} = z'_{i_1}, \ldots, z_{i_{p_+}} = z'_{i_{p_+}}.$$

Hence $\mathbf{v} = \mathbf{v}'$ and thus $u = u'$. Thus f is injective on $\mathcal{L}$. Finally, from the definition of $\mathcal{L}$ and $\mathcal{L}'$, one easily verifies that, for every $u \in \mathcal{L}$, $\psi(u) = \psi(f(u))$. This concludes the proof. □

Proof of Lemma 9: By contradiction, assume that there exist two languages $\mathcal{L}'_i = L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ and $\mathcal{L}'_j = L'(\mathbf{j}, \delta_1, \ldots, \delta_{k_j}, e_1, \ldots, e_{k_j})$ such that $\mathcal{L}'_i \cap \mathcal{L}'_j \neq \emptyset$. Let us denote by $\mathcal{B}_i$ the set $B(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ associated with the language $\mathcal{L}'_i$. By (37), $\mathcal{B}_i$ is the set:

$$\{\mathbf{b}^{(i)} + \sum_{\ell=1}^{p_a} x_{i_\ell} N_{a,i_\ell} \mathbf{b}^{(i)}_{a,i_\ell} + \sum_{\ell=1}^{p_c} y_{i_\ell} N_{c,i_\ell} \mathbf{b}^{(i)}_{c,i_\ell} + \sum_{\ell=1}^{p_+} z_{i_\ell} N_{+,i_\ell} \mathbf{b}^{(i)}_{+,i_\ell} \mid x_{i_\ell},\ y_{i_\ell},\ z_{i_\ell} \geq 1\}.$$

Similarly, denote $\mathcal{B}_j$ the set $B(\mathbf{j}, \delta_1, \ldots, \delta_{k_j}, e_1, \ldots, e_{k_j})$ associated with the language $\mathcal{L}'_j$. By (37), $\mathcal{B}_j$ is the set:

$$\{\mathbf{b}^{(j)}+\sum_{\ell=1}^{p'_a} x_{i_\ell} N_{a,j_\ell}\mathbf{b}^{(j)}_{a,j_\ell}+\sum_{\ell=1}^{p'_c} y_{j_\ell} N_{c,j_\ell}\mathbf{b}^{(j)}_{c,j_\ell}+\sum_{\ell=1}^{p'_+} z_{j_\ell} N_{+,j_\ell}\mathbf{b}^{(j)}_{+,j_\ell} \mid x_{j_\ell},\ y_{j_\ell},\ z_{j_\ell} \geq 1\}.$$

By hypothesis, there exists $u \in \mathcal{L}'_i \cap \mathcal{L}'_j$. We want to prove that $\mathcal{B}_i = \mathcal{B}_j$, which is a contradiction. Since $u \in \mathcal{L}'_i$, by (39), we have:

$$u = \varphi(\widetilde{\mathbf{b}}^{(i)}) w_{i_1}^{(k_i - n_a - n_c)} \mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_{i_1}} w_{i_2}^{z_{i_2}} \cdots w_{i_{p_+}}^{z_{i_{p_+}}}, \tag{62}$$

where

- $\widetilde{\mathbf{b}}^{(i)} = \mathbf{b}_0^{(i)} \ + \ (d_{i_1} - N_{i_1}^{(i)} k_i)\mathbf{b}_{i_1}^{(i)} \ + \ \sum_{\ell=1,\ \ell\neq i_1}^{k_i} \ d_\ell \mathbf{b}_\ell^{(i)}$,
- $\mathrm{u}_a \in \mathrm{L}_a^{(i)},\ \mathrm{u}_c \in \mathrm{L}_c^{(i)}$,
- $z_{i_1}, \ldots, z_{i_{p_+}}$, is a tuple of integers ≥ 1.

Similarly, since $u \in \mathcal{L}'_j$, by (39), we have:

$$u = \varphi(\widetilde{\mathbf{b}}^{(j)}) w_{j_1}^{(k_j - n'_a - n'_c)} \mathrm{u}'_a \mathrm{u}'_c w_{j_1}^{z'_{j_1}} w_{j_2}^{z'_{j_2}} \cdots w_{j_{p'_+}}^{z'_{i_{p'_+}}}, \tag{63}$$

where

- $\widetilde{\mathbf{b}}^{(j)} = \mathbf{b}_0^{(j)} \ + \ (e_{j_1} - N_{j_1}^{(j)} k_j)\mathbf{b}_{j_1}^{(j)} \ + \ \sum_{\ell=1,\ \ell\neq j_1}^{k_j} \ e_\ell \mathbf{b}_\ell^{(j)}$,
- $\mathrm{u}'_a \in \mathrm{L}_a^{(j)}, \mathrm{u}'_c \in \mathrm{L}_c^{(j)}$,
- $z'_{i_1}, \ldots, z'_{i_{p'_+}}$, is a tuple of integers ≥ 1.

Let

$$U_i = w_{i_1}^{(k_i - n_a - n_c)} \mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_{i_1}} w_{i_2}^{z_{i_2}} \cdots w_{i_{n_+}}^{z_{i_{p_+}}},$$

and

$$U_j = w_{j_1}^{(k_j - n'_a - n'_c)} \mathrm{u}'_a \mathrm{u}'_c w_{j_1}^{z'_{j_1}} w_{j_2}^{z'_{j_2}} \cdots w_{j_{p'_+}}^{z'_{i_{p'_+}}}.$$

Let us first show that $|U_i| = |U_j|$. Indeed, assume $|U_i| < |U_j|$ (the other case is treated similarly). Since all the words appearing in U_i and U_j have the same length β, the latter implies that w_{i_1} is a suffix of $\varphi(\widetilde{\mathbf{b}}_j)$. Since $\varphi(\widetilde{\mathbf{b}}_j) \in u_1^* \cdots u_k^*$ and since (cf Remark 2) w_{i_1} cannot be a factor of any word of $u_1^* \cdots u_k^*$, one gets a contradiction. Hence $|U_i| = |U_j|$. The latter implies that

$$U_i = U_j, \quad \varphi(\widetilde{\mathbf{b}}_i) = \varphi(\widetilde{\mathbf{b}}_j).$$

From $\varphi(\widetilde{\mathbf{b}}_i) = \varphi(\widetilde{\mathbf{b}}_j)$, since $\widetilde{\mathbf{b}}_i, \widetilde{\mathbf{b}}_j \in B$ and φ is injective on B, one has

$$\widetilde{\mathbf{b}}_i = \widetilde{\mathbf{b}}_j. \tag{64}$$

This implies that $\mathcal{B}_i$ and $\mathcal{B}_j$ come from the same simple set B_i of the partition of B, and hence we get

$$k_i = k_j,\ n_a = n'_a,\ n_c = n'_c,$$

so yielding $w_{i_1}^{(k_i - n_a - n_c)} = w_{j_1}^{(k_j - n'_a - n'_c)}$. From the latter, since all the words appearing in U_i and U_j have the same length β and all the words of w_{i_ℓ} contain at least two distinct letters, one has

$$\mathrm{u}_a = \mathrm{u}'_a, \tag{65}$$

$$\mathrm{u}_c = \mathrm{u}'_c, \tag{66}$$

and

$$w_{i_1}^{z_{i_1}} w_{i_2}^{z_{i_2}} \cdots w_{i_{p_+}}^{z_{i_{p_+}}} = w_{j_1}^{z'_{j_1}} w_{j_2}^{z'_{j_2}} \cdots w_{j_{p'_+}}^{z'_{i_{p'_+}}}. \tag{67}$$

From (67), since we are dealing with code words, one has $p_+ = p'_+$ and, for every $\ell = 1, \ldots, p_+$, $w_{i_\ell} = w_{j_\ell}$ and $z_{i_\ell} = z'_{j_\ell}$. Hence, by using the coding (23) of Section 4.3, one has:

$$\forall\ \ell = 1, \ldots, p_+ \quad N_{i_\ell}^{(i)} \mathbf{b}_{+,i_\ell}^{(i)} = N_{j_\ell}^{(j)} \mathbf{b}_{+,j_\ell}^{(j)}. \tag{68}$$

From (65) and $n_a = n'_a$, one has $p_a = p'_a$ and, by Remark 5, one has:

$$\forall\ \ell = 1, \ldots, p_a \quad N_{i_\ell}^{(i)} \mathbf{b}_{a,i_\ell}^{(i)} = N_{j_\ell}^{(j)} \mathbf{b}_{a,j_\ell}^{(j)}. \tag{69}$$

Similarly, from (66) and $n_c = n'_c$, one has $p_c = p'_c$ and,

$$\forall\ \ell = 1, \ldots, p_c \quad y_{i_\ell} N_{i_\ell}^{(i)} \mathbf{b}_{c,i_\ell}^{(i)} = y'_{j_\ell} N_{j_\ell}^{(j)} \mathbf{b}_{c,j_\ell}^{(j)}. \tag{70}$$

Finally, from (64), (68), (69), and (70), it follows that $\mathcal{B}_i = \mathcal{B}_j$, a contradiction. The proof of the lemma is complete. □

Proof of Lemma 17: Let us prove that $L'_+ \cap L'_{ac} \neq \emptyset$. By contradiction, assume that there exists a word u with $u \in L'_+ \cap L'_{ac}$. Since $u \in L'_+$, by the definition of L'_+, there exists a language $\mathcal{L}'_i = L'(\mathbf{i}, \epsilon_1, \ldots, \epsilon_{k_i}, d_1, \ldots, d_{k_i})$ defined by (39) such that $u \in \mathcal{L}'_i$. Hence, u is written in the form:

$u = \varphi(\widetilde{\mathbf{b}}^{(i)}) w_{i_1}^{(k_i-n_a-n_c)} \mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_{i_1}} w_{i_2}^{z_{i_2}} \cdots w_{i_{p_+}}^{z_{ip_+}}$, where:

- $\widetilde{\mathbf{b}}^{(i)} = \mathbf{b}_0^{(i)} \; + \; (d_{i_1} - N_{i_1}^{(i)} k_i)\mathbf{b}_{i_1}^{(i)} \; + \; \sum_{\ell=1,\ \ell\neq i_1}^{k_i} \; d_\ell \mathbf{b}_\ell^{(i)}$,
- $\mathrm{u}_a \in \mathrm{L}_a^{(i)}, \; \mathrm{u}_c \in \mathrm{L}_c^{(i)}$,
- $z_{i_1}, \ldots, z_{i_{p_+}}$, is a tuple of integers ≥ 1.

Similarly, since $u \in L'_{ac}$, by the definition of L'_{ac}, there exists a language $\mathcal{L}'_j = L'(\mathbf{j}, \delta_1, \ldots, \delta_{k_j}, e_1, \ldots, e_{k_j})$ defined by (49) such that $u \in \mathcal{L}'_j$. Hence, u is written in the form:

$u = \varphi(\widetilde{\mathbf{b}}^{(j)}) w^{(k_j-m_a-m_c)} \mathrm{u}'_a \mathrm{u}'_c w$, where:

- $\widetilde{\mathbf{b}}^{(j)} = \mathbf{b}_0^{(j)} + (e_{j_1} - N_{j_1}^{(j)}(k_j+1))\mathbf{b}_{j_1}^{(j)} + (e_{j_2} - N_{j_2}^{(j)}(k_j+1))\mathbf{b}_{j_2}^{(j)} + \sum_{\ell=1,\ \ell\neq j_1,j_2}^{k_j} \; e_\ell \mathbf{b}_\ell^{(j)}$.
- $\mathrm{u}'_a \in \mathrm{L}_a^{(j)}, \; \mathrm{u}'_c \in \mathrm{L}_c^{(j)}$,
- w satisfies Property 1 of Section 4.6.1.

By comparing the latter two factorizations of u, and by using the very same argument of the proof of Lemma 9, one gets $\varphi(\widetilde{\mathbf{b}}^{(i)}) = \varphi(\widetilde{\mathbf{b}}^{(j)})$, $k_i = k_j, n_a = m_a$, and $n_c = m_c$. Thus one obtains the equation

$w_{i_1}^{(k_i-n_a-n_c)} \mathrm{u}_a \mathrm{u}_c w_{i_1}^{z_{i_1}} w_{i_2}^{z_{i_2}} \cdots w_{i_{p_+}}^{z_{ip_+}} = w^{(k_i-n_a-n_c)} \mathrm{u}'_a \mathrm{u}'_c w$.

Since $|w| = 2\beta$ and $|w_{i_1}| = \beta$, either $w = w_{i_1}^2$ or $w = w_{i_1} a^\beta$. By Property 1 of Section 4.6.1, the latter two factorizations are not possible for w. Hence $L'_+ \cap L'_{ac} = \emptyset$.

Let us prove that $L'_a \cap L'_c = \emptyset$. By Theorem 6 and Theorem 7, the regular languages L'_a and L'_c are commutatively equivalent to L_a and L_c, respectively. The claim follows by applying Corollary 2.

Now let us prove that, for every $\sigma \in A, L'_\sigma \cap L'_+ = \emptyset$. By contradiction, assume that there exists a word $v \in L'_+ \cap L'_\sigma$, with $\sigma \in A$. Since $v \in L'_+$, by Theorem 4, v has a suffix of length β that cannot be a factor of any word of $u_1^* \cdots u_k^*$. On the opposite, since $v \in L'_\sigma$, by Theorem 6 or Theorem 7, v has a suffix of length 2β which is a factor of a word of $u_1^* \cdots u_k^*$. This implies the result. The same argument shows that, for every $\sigma \in A, L'_\sigma \cap L'_{ac} = \emptyset$. □

4.9 The proof of Proposition 5

Recall that, by (57) and (58), for every $\ell = 1, \ldots, s_u$, $L^a_{t,u,\ell} = \varphi(B^a_{t,u,\ell})$. Consider any word w in $L^a_{t,u,\ell}$. With every occurrence of the letter a in the word w we assign a positive integer $\ell \leq t+1$, called *a-index ℓ in w*, defined as follows:

- if there exists an r, with $1 \leq r \leq t-1$, such that the position in w of the occurrence of a is on the right of the r-th occurrence of c and on the left of the of $r+1$-th occurrence of c, then we set $\ell = r+1$;
- if the position of the occurrence of a is on the right of the t-th occurrence of c, then we set $\ell = t+1$;
- if the position of the occurrence of a is on the left of the first occurrence of c, then we set $\ell = 1$.

Since the proof of Proposition 5 is rather technical, Example 4 of the Appendix will clarify all the steps of the proof. For the sake of simplicity, denote by B a set $B^a_{t,u,\ell}$ of the decomposition (58). Assume that $B = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m : x_\ell \geq 1,\ 1 \leq \ell \leq m\}$. Now let $x_1, \ldots, x_m \geq 1$ be given, and let the word w defined by $w = \varphi(\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m)$, where $\varphi : \mathbb{N}^k \longrightarrow u_1^* \cdots u_k^*$ is the Ginsburg map defined in (2). Observe that by the definition of the Ginsburg map:

$$w = u_1^{b_{01}+\sum_{h=1}^m x_h b_{h1}} \cdots u_j^{b_{0j}+\sum_{h=1}^m x_h b_{hj}} \cdots u_k^{b_{0k}+\sum_{h=1}^m x_h b_{hk}}.$$

Now for every $j = 1, \ldots, k$, we have $u_j^{b_{0j}+\sum_{h=1}^m x_h b_{hj}} = u_j^{b_{0j}} u_j^{x_1 b_{1j}} \cdots u_j^{x_m b_{mj}}$. Given a vector $\mathbf{b}_i$, with $1 \leq i \leq m$, of the representation of B, we say that $u_j^{x_i b_{ij}}$ is *the factor of w corresponding to the j-th component of $x_i\mathbf{b}_i$*.

Lemma 18 *Let $\mathbf{b}_i = (b_{i1}, \ldots, b_{ik})$ with $1 \leq i \leq m$. For every non-zero component b_{ij} of $\mathbf{b}_i$, with $1 \leq j \leq k$, there exists an r, with $1 \leq r \leq t+1$ such that, for every $x_1, \ldots, x_m \geq 1$, if $w = \varphi(\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m)$, then the a-index in w of every occurrence of a in the factor of w corresponding to the j-th component of the vector $x_i\mathbf{b}_i$ has always the value r.*

Proof. Let $\mathbf{b}_i = (b_{i1}, \ldots, b_{ik})$, with $1 \leq i \leq m$, and a non-zero component b_{ij} of $\mathbf{b}_i$, with $1 \leq j \leq k$ be fixed. Then either there exists a greatest index j_0, with $1 \leq j_0 \leq k$ such that:

– $|u_{j_0}|_c > 0$, and $b_{0j_0} \neq 0$,
– $j > j_0$,

or no such index exists.
Then we claim that:

– in the first case the required value r is $\sum_{h=1}^{j_0} |u_h|_c \, b_{0h} + 1$,
– in the second case the required value r is 1.

Let $w = \varphi(\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m)$, with $x_1, \ldots, x_m \geq 1$. Then, in case the index j_0 exists, the last c occurring in w before any occurrence of a in $u_j^{x_i b_{ij}}$ is the last occurrence of c in $u_{j_0}^{b_0 j_0}$ and one immediately checks that such occurrence is the $(r-1)$-th occurrence of c. In the second case there are no occurrences of c on the left of $u_j^{x_i b_{ij}}$ in w as no component of $\mathbf{b}_0$ generates it. □

As shown by the previous lemma, for every vector $\mathbf{b}_i$, with $1 \leq i \leq m$, for every component b_{ij} of $\mathbf{b}_i$, with $1 \leq j \leq k$, and for every $x_i \in \mathbb{N}$, with $x_i > 0$, the a-index in w of every occurrence of a in $u_j^{x_i b_{ij}}$ has a value which depends only on the component b_{ij}. Therefore, by a minor terminological abuse, we call this value *the a-index of the component b_{ij}*.

Proof of Proposition 5: Let $\ell \in \mathbb{N}$ be fixed. For every vector $\mathbf{b}_i$, with $1 \leq i \leq m$, we let $\mathcal{T}_{i,\ell}$ be the set of all indexes h, with $1 \leq h \leq k$, such that the component b_{ih} has a-index ℓ. Now we define the vector $\widehat{\mathbf{b}}_i$ as follows:

$$\widehat{\mathbf{b}}_i = (\widehat{b}_{i1}, \ldots, \widehat{b}_{it+1}),$$

where, for $\ell = 1, \ldots, t+1$, $\widehat{b}_{i\ell} = \sum_{h \in \mathcal{T}_{i,\ell}} b_{ih}|u_h|$. We define the vector $\widehat{\mathbf{b}}_0$ as:

$$\widehat{\mathbf{b}}_0 = (\widehat{b}_{01}, \ldots, \widehat{b}_{0t+1}),$$

where, for $\ell = 1, \ldots, t+1$, $\widehat{b}_{0\ell}$ is the number of the occurrences of a in $\varphi(\mathbf{b}_0)$ with a-index ℓ in $\varphi(\mathbf{b}_0)$.

Finally we define $\widehat{B}$ as $\widehat{B} = \{\widehat{\mathbf{b}}_0 + x_1\widehat{\mathbf{b}}_1 + \cdots + x_m\widehat{\mathbf{b}}_m : x_1, \ldots, x_m \geq 1\}$. Now we want to prove the following claim:

Claim. $\widehat{\varphi}(\widehat{\mathbf{b}}_0 + x_1\widehat{\mathbf{b}}_1 + \cdots + x_m\widehat{\mathbf{b}}_m) = \varphi(\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m)$, for every $x_1, \ldots, x_m \geq 1$.

Let us denote by w the word $\varphi(\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m)$. To prove the claim, consider any vector $x_i\mathbf{b}_i$, with $1 \leq i \leq m$, with $x_i\mathbf{b}_i = (x_ib_{i1}, \ldots, x_ib_{ik})$. Let the a-index ℓ be fixed, with $1 \leq \ell \leq t+1$. By Lemma 18, for every component b_{ih} with a-index ℓ, the word $u_i^{x_i b_{ih}}$ contributes a string of a of length $x_i b_{ih}|u_h|$ to its a-index. Therefore the contribution of vector $x_i\mathbf{b}_i$ to the a-index ℓ in w is given by $\sum_{h\in\mathcal{T}_{i,\ell}} x_i b_{ih}|u_h|$. On the other hand, $\widehat{b}_{i\ell}$ is by definition equal to $\sum_{h\in\mathcal{T}_{i,\ell}} b_{ih}|u_h|$ and therefore $x_i\widehat{b}_{i\ell}$ contributes $\sum_{h\in\mathcal{T}_{i,\ell}} x_i b_{ih}|u_h|$ to the a-index ℓ. The claim follows. $\diamond$

From the claim it is immediate that $\widehat{\varphi}(\widehat{B}) = \varphi(B)$, and since φ is injective on B then $\widehat{\varphi}$ is injective on $\widehat{B}$. It is moreover clear that $\widehat{B}$ is a simple set.

Let us prove the second part of the statement of the proposition. We first do the following remarks on the vectors $\widehat{\mathbf{b}}_i$ defined above:

1) for every $i = 1,\dots,m$, and for every $j = m_u+2,\dots,t+1$, the j-th-component of the vector $\widehat{\mathbf{b}}_i$ is null. This immediately follows from that fact that $\widehat{\varphi}(\widehat{B}^a_{t,u,\ell})$ is a subset of the product $\underbrace{a^*ca^*ca^*c\cdots ca^*}_{m_u-times}u$, where m_u is the number occurrences of the letter c in the left side part of such product and $t = |u|_c + m_u$. For every $i = 1,\dots,m$, let us define the vectors of $\mathbb{N}^{m_u+1}$ as $\widehat{\widehat{\mathbf{b}}}_i = (\widehat{b}_{i1},\dots,\widehat{b}_{im_u+1})$. Observe that $\widehat{\mathbf{b}}_i = (\widehat{\widehat{\mathbf{b}}}_i, \underbrace{0,0,\dots,0}_{t-m_u})$, and that the vectors $\widehat{\widehat{\mathbf{b}}}_i$, with $i = 1,\dots,m$, are obviously linearly independent in $\mathbb{N}^{m_u+1}$.

2) let $u = a^{\alpha_1}ca^{\alpha_2}c\cdots ca^{\alpha_n}ca^{\alpha_{n+1}}$ be the factorization of u, with $n = |u|_c$ and $\alpha_1,\dots,\alpha_{n+1} \geq 0$. Since $L^a_{t,u,\ell} = \widehat{\varphi}(\widehat{B})$ and the previous point (1), one has that, in the vector $\widehat{\mathbf{b}}_0$, $\widehat{b}_{0m_u+1} \geq \alpha_{m_u+1}$ and, for every $j = m_u+2,\dots,t+1$, $\widehat{b}_{0j} = \alpha_j$.

Let $\widehat{\widehat{\mathbf{b}}}_0$ be the vector of $\mathbb{N}^{m_u+1}$ defined as $\widehat{\widehat{\mathbf{b}}}_0 = (\widehat{b}_{01},\dots,\widehat{b}_{0m_u},\widehat{b}_{0m_u+1}-\alpha_1)$, and define the vector $\mathbf{b}_u$ of $\mathbb{N}^{t+1}$ as

$$\mathbf{b}_u = (\underbrace{0,0,\dots,0}_{m_u},\alpha_1,\dots,\alpha_{n+1}). \tag{71}$$

Observe that the definition of the vector $\mathbf{b}_u$ depends only from u and not from the simple set B. Moreover, observe that $\widehat{\mathbf{b}}_0 = (\widehat{\widehat{\mathbf{b}}}_0, \underbrace{0,0,\dots,0}_{t-m_u}) + \mathbf{b}_u$.

Set $\widehat{D} = \{\widehat{\widehat{\mathbf{b}}}_0 + x_1\widehat{\widehat{\mathbf{b}}}_1 + \cdots + x_m\widehat{\widehat{\mathbf{b}}}_m : x_1,\dots,x_m \geq 1\}$. By construction, $\widehat{D}$ is a simple set of $\mathbb{N}^{m_u+1}$ and $\widehat{B} = \widehat{D} \times 0^{t-m_u} + \mathbf{b}_u$. The proof is complete. $\square$

Remark 8 The reader may wonder how Proposition 5 can be proved without showing that the number m of the generators of B is no greater than $t+1$. Indeed the bound $m \leq t+1$ is implicit in the hypothesis that φ is injective on B. However, such bound can be proved in a direct way.

Extension of Theorem 1

We now sketch the proof of the extension of Theorem 1 to a general alphabet $A = \{a_1, \ldots, a_s\}$ with $s \geq 3$ letters. Let L be a bounded language of A^* such that $L = \varphi(B)$, where B is a semi-simple set. By using the algorithm of Section 4.4, we construct a partition of semi-simple sets of B:

$$B = B_- \cup B_+ \cup C \cup \bigcup_{i=1}^{s} B_{a_i}, \qquad (1)$$

where the sets of (1) are defined as follows:

1. B_- is a finite set of vectors;

2. B_+ is a finite union of pairwise disjoint simple sets of dimension ≥ 1, every one of each satisfies the following property: if $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of the simple set, there exists ℓ with $1 \leq \ell \leq m$ where $\varphi(\mathbf{b}_\ell)$ contains at least two distinct letters;

3. C is a finite union of pairwise disjoint simple sets of dimension ≥ 1, every one of each satisfies the following property. Let $a_{i_1}, \ldots, a_{i_\ell}$ be distincts letters with $\ell \geq 2$. If $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of the simple set, for every $\ell = 1, \ldots, m$, $\varphi(\mathbf{b}_\ell)$ is a non-trivial power of some a_{i_j}; moreover, no index ℓ exists such that $1 \leq \ell \leq m$ and $\varphi(\mathbf{b}_\ell)$ contains at least two distinct letters;

4. Let a_i be a letter of A. Then B_{a_i} is a finite union of pairwise disjoint simple sets of dimension ≥ 1, every one of each satisfies the following property: if $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ form the representation of the simple set, for every $\ell = 1, \ldots, m$, $\varphi(\mathbf{b}_\ell) \in a_i^+$.

Finally, let L_-, L_+, L_C and, for every $i = 1, \ldots, s$, L_{a_i} be the image under the map φ of the sets of the decomposition (1). Then one has:

$$L = L_- \cup L_+ \cup L_C \cup \bigcup_{i=1}^{s} L_{a_i}. \qquad (2)$$

Moreover, up to a slight refinement of the decomposition (2), we may suppose that, for every i, j with $1 \leq i \neq j \leq s$, one has:

$$\forall\, u \in L_{a_i},\ \forall\, v \in L_{a_j}, |u|_{a_j} < |v|_{a_j}. \qquad (3)$$

Now we describe the construction of the regular languages which are commutatively equivalent to the languages of the decomposition (2). By using the technique of Section 4.5, we construct a regular language L'_+ such that $L'_+ \sim L_+$. Similarly, by using the technique of Section 4.7, for every letter a_i, with $1 \leq i \leq s$, we construct a regular language L'_{a_i} such that $L'_{a_i} \sim L_{a_i}$.

Let us describe the construction of the regular language L'_C commutatively equivalent to L_C. For this purpose, let $\mathcal{B}$ be an arbitrary simple set of the decomposition of C and let $\mathbf{b}_0, \mathbf{b}_1, \ldots, \mathbf{b}_m$ be the vectors of its representation. By hypothesis, we may suppose that $\varphi(\mathbf{b}_1), \ldots, \varphi(\mathbf{b}_{n_1}) \in \sigma_1^+, \ldots, \varphi(\mathbf{b}_j), \ldots, \varphi(\mathbf{b}_{n_\ell}) \in \sigma_\ell^+$, where $\sigma_1, \ldots, \sigma_\ell$ are distinct letters of A, with $\ell \geq 2$. In the case of two letters σ_1, σ_2, we associate with $\mathcal{B}$ the regular language

$$\varphi(\widetilde{\mathbf{b}}^{(i)}) w^{(k_i - n_{\sigma_1} - n_{\sigma_2})} \mathrm{L}^{(i)}_{\sigma_1} \mathrm{L}^{(i)}_{\sigma_2} w, \qquad (4)$$

of the Eq. (49) where $\mathrm{L}^{(i)}_{\sigma_1}$ and $\mathrm{L}^{(i)}_{\sigma_2}$ are defined by the Eq. (47). In the case of three or more letters, the regular language associated with $\mathcal{B}$ is defined similarly as:

$$\varphi(\widetilde{\mathbf{b}}^{(i)}) w^{(k_i - (n_{\sigma_1} + n_{\sigma_2} + \cdots + n_{\sigma_\ell}))} \mathrm{L}^{(i)}_{\sigma_1} \mathrm{L}^{(i)}_{\sigma_2} \mathrm{L}^{(i)}_{\sigma_3} \cdots \mathrm{L}^{(i)}_{\sigma_\ell} w, \qquad (5)$$

where, for every $j = 1, \ldots, \ell$, the language $\mathrm{L}^{(i)}_{\sigma_j}$ is defined as in (47) and it codifies the vectors of the representation of $\mathcal{B}$ whose images, under the map φ, are powers of the letter σ_j. By following the very same argument of Lemma 11, one can prove that the languages (4) and (5) are pairwise disjoint. Finally the regular language L'_C is defined as the union of all the languages (4) and (5) associated with the simple sets $\mathcal{B}$ of the decomposition of C. Let us define the regular language L' as

$$L' = L'_+ \ \cup \ L'_C \ \cup \ \textstyle\bigcup_{i=1}^{s} L'_{a_i}. \qquad (6)$$

The languages of the decomposition (6) of L' are pairwise disjoint. Indeed, by (3) and by Corollary 2, the languages L'_{a_i}, with $1 \leq i \leq s$, are pairwise disjoint. Moreover, since the definitions of the languages of the decomposition (6) of L' are exactly the same given in Section 4, the same argument of Lemma 17 allows to prove the claim.
Hence, by Lemma 1, L' is commutatively equivalent to $L_+ \cup L_C \cup \bigcup_{i=1}^{s} L_{a_i}$. By Lemma 2, there exists a finite set of words L'_- such that the languages L and $L'_- \cup L'$ are commutatively equivalent. This completes the proof.

Examples

Example 1 Let $B = \{x_1\mathbf{b}_1 + x_2\mathbf{b}_2 : x_1, x_2 \in \mathbb{N}\}$ be a simple set of $\mathbb{N}^k, k \geq 2$, where $\mathbf{b}_0 = \mathbf{0}$, $\mathbf{b}_1$ and $\mathbf{b}_2$ form the representation of B. Let $N_1 = 2,\ N_2 = 3$ and $\chi_1 = \chi_2 = 0$. One has $B = B^- \cup B^{(-,+)} \cup B^{(+,-)} \cup B^{(+,+)}$, where:
$B^- = \{c_1\mathbf{b}_1 + c_2\mathbf{b}_2 : 0 \leq c_1 < 2,\ 0 \leq c_2 < 3\}$,
$B^{(-,+)} = \{c_1\mathbf{b}_1 + x_2\mathbf{b}_2 : 0 \leq c_1 < 2,\ x_2 \geq 3\}$,
$B^{(+,-)} = \{x_1\mathbf{b}_1 + c_2\mathbf{b}_2 : 0 \leq c_2 < 3,\ x_1 \geq 2\}$,
$B^{(+,+)} = \{x_1\mathbf{b}_1 + x_2\mathbf{b}_2 : x_1 \geq 2, x_2 \geq 3\}$.
For $(\epsilon_1, \epsilon_2) = (+,+)$, $B^{(+,+)}$ is partitioned into the family of 6 simple sets of the form $B(+,+,r_1,r_2) = \{(r_1\mathbf{b}_1 + r_2\mathbf{b}_2) + 2x_1\mathbf{b}_1 + 3x_2\mathbf{b}_2 : x_1, x_2 \geq 1\}$, where $0 \leq r_1 < 2,\ 0 \leq r_2 < 3$.

For $(\epsilon_1, \epsilon_2) = (-,+)$, $B^{(-,+)}$ is partitioned into the family given by 3 simple sets of the form $B(-,+,0,r_2) = \{r_2\mathbf{b}_2 + 3x_2\mathbf{b}_2 : x_2 \geq 1\}$, together with 3 sets of the form $B(-,+,1,r_2) = \{\mathbf{b}_1 + r_2\mathbf{b}_2 + 3x_2\mathbf{b}_2 : x_2 \geq 1\}$, where $0 \leq r_2 < 3$.

Example 2 Let us consider the simple set B_i of dimension $k_i = 8$ whose representation is given by the vectors: $\mathbf{b}_0,\ \mathbf{b}_1,\ \mathbf{b}_2,\ \mathbf{b}_3,\ \mathbf{b}_4,\ \mathbf{b}_5,\ \mathbf{b}_6,\ \mathbf{b}_7,\ \mathbf{b}_8$. We suppose that:
$\forall\ \ell = 1, \ldots, 4,\ \varphi(\mathbf{b}_\ell) \in a^+$;
$\forall\ \ell = 5, 6,\ \varphi(\mathbf{b}_\ell) \in c^+$;
$\forall\ \ell = 7, 8,\ \varphi(\mathbf{b}_\ell)$ contains at least two distinct letters.

Thus the list above is written as: $\mathbf{b}_0, \mathbf{b}_{a,1}, \mathbf{b}_{a,2}, \mathbf{b}_{a,3}, \mathbf{b}_{a,4}, \mathbf{b}_{c,1}, \mathbf{b}_{c,2}, \mathbf{b}_{+,1}, \mathbf{b}_{+,2}$, so that $n_a = 4,\ n_c = 2,\ n_+ = 2$.

Let $\mathcal{B} = B(\mathbf{i}, +, +, +, +, -, -, +, +, d_1, d_2, d_3, d_4, d_5, d_6, d_7, d_8)$ be the simple set of the decomposition of B_+ defined in (37) and given by all the vectors:

$$\mathbf{b} + x_1N_{a,1}\mathbf{b}_{a,1} + x_2N_{a,2}\mathbf{b}_{a,2} + x_3N_{a,3}\mathbf{b}_{a,3} + x_4N_{a,4}\mathbf{b}_{a,4} + z_1N_{+,1}\mathbf{b}_{+,1} + z_2N_{+,2}\mathbf{b}_{+,2},$$

where $x_\ell, z_\ell \geq 1$ and $\mathbf{b} = \mathbf{b}_0 + \sum_{\ell=1}^{8} d_\ell\mathbf{b}_\ell$. We can rewrite $\mathbf{b}$ as:

$$\mathbf{b}_0 + d_{a,1}\mathbf{b}_{a,1} + d_{a,2}\mathbf{b}_{a,2} + d_{a,3}\mathbf{b}_{a,3} + d_{a,4}\mathbf{b}_{a,4} + d_{c,1}\mathbf{b}_{c,1} + d_{c,2}\mathbf{b}_{c,2} + d_{+,1}\mathbf{b}_{+,1} + d_{+,2}\mathbf{b}_{+,2}.$$

Observe that, by the definition of $\mathcal{B}$, $d_{+,1} \geq 9N_{+,1}\beta$. The regular language associated with $\mathcal{B}$ is $\varphi(\widetilde{\mathbf{b}})w_1^2\mathrm{L}_a^{(i)}\mathrm{L}_c^{(i)}w_1^+w_2^+$ (cf Eq. (39)) where:
– w_1 and w_2 are the words of $\mathcal{W}$ associated with $N_{+,1}\mathbf{b}_{+,1}, N_{+,2}\mathbf{b}_{+,2}$ respectively;
– $\mathrm{L}_a^{(i)} = w_1(a^\beta)^+w_1(a^\beta)^+w_1(a^\beta)^+w_1(a^\beta)^+$;
– $\mathrm{L}_c^{(i)} = w_1^2$;
– $\widetilde{\mathbf{b}}$ is the vector obtained from $\mathbf{b}$ by replacing $d_{+,1}$ with $(d_{+,1} - 8N_{+,1})$.

Example 3 Let B_i be the simple set of Example 2. Consider the simple set $\mathcal{B} = B(\mathbf{i}, -, +, -, +, +, +, -, -, e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8)$ of the decomposition of $B_{a,c}$ defined as in (46):

$$\{\mathbf{b} + x_2 N_{a,2}\mathbf{b}_{a,2} + x_4 N_{a,4}\mathbf{b}_{a,4} + y_1 N_{c,1}\mathbf{b}_{c,1} + y_2 N_{c,2}\mathbf{b}_{c,2} \mid x_\ell,\ y_\ell \geq 1\},$$

where $\mathbf{b} = \mathbf{b}_0 \ + \ \sum_{\ell=1}^{8} \ e_\ell \mathbf{b}_\ell$. We can rewrite $\mathbf{b}$ as:

$$\mathbf{b}_0 + e_{a,1}\mathbf{b}_{a,1} + e_{a,2}\mathbf{b}_{a,2} + e_{a,3}\mathbf{b}_{a,3} + e_{a,4}\mathbf{b}_{a,4} + e_{c,1}\mathbf{b}_{c,1} + e_{c,2}\mathbf{b}_{c,2} + e_{+,1}\mathbf{b}_{+,1} + e_{+,2}\mathbf{b}_{+,2}.$$

Observe that, by the definition of $\mathcal{B}$, $e_{a,2} \geq 9N_{a,2}\beta$ and $e_{c,1} \geq 9N_{c,1}\beta$. Then the regular language associated with $\mathcal{B}$ is $\varphi(\widetilde{\mathbf{b}})w^2\mathrm{L}_a^{(i)}\mathrm{L}_c^{(i)}w$, where:
– w is a word of A^* satisfying Property 1 of Section 4.6;
– $\mathrm{L}_a^{(i)} = ww(a^\beta)^+ww(a^\beta)^+$;
– $\mathrm{L}_c^{(i)} = w(c^\beta)^+w(c^\beta)^+$;
– $\widetilde{\mathbf{b}}$ is the vector obtained from $\mathbf{b}$ by replacing $e_{a,2}$ with $(e_{a,2} - 9N_{a,2})$ and $e_{c,1}$ with $(e_{c,1} - 9N_{c,1})$, respectively.

Example 4 Assume $k = 7$ and let $u_1 = a^2$, $u_2 = a\mathbf{c}aa$, $u_3 = a$, $u_4 = \mathbf{c}$, $u_5 = a^7$, $u_6 = \mathbf{c}$, and $u_7 = a^2$. We find useful to emphasize in **bold** the occurrences of the symbol c in the factorizations of words. In the word $u_1u_2^2u_3u_4u_5u_6u_7 = aaa\mathbf{c}aaa\mathbf{c}aaa\mathbf{c}a^7\mathbf{c}a^2$, for instance, the first 3 occurrences of a have a-index 1, the subsequent 3 occurrences of a have a-index 2, while the last 2 occurrences of a have a-index 5. Let us consider the simple set of $\mathbb{N}^7$ $B = \mathbf{b}_0 + \{\mathbf{b}_1, \mathbf{b}_2\}^\oplus$ where $\mathbf{b}_0 = (0, 2, 0, 1, 7, 1, 2)$, $\mathbf{b}_1 = (2, 0, 0, 0, 0, 0, 0)$, and $\mathbf{b}_2 = (2, 0, 1, 0, 0, 0, 0)$. Then, for every $x, y \in \mathbb{N}$, $\mathbf{b}_0 + x\mathbf{b}_1 + y\mathbf{b}_2 = (2x + 2y, 2, y, 1, 7, 1, 2)$, so that $\varphi(\mathbf{b}_0 + x\mathbf{b}_1 + y\mathbf{b}_2) = u_1^{2x+2y}u_2^2u_3^yu_4u_5^7u_6u_7^2 = a^{4x+4y}a\mathbf{c}aaa\mathbf{c}aaa^y\mathbf{c}a^7\mathbf{c}a^2$. Observe that, with respect to the vector $\mathbf{b}_2$, the a-index of every occurrence of a in $u_1^{yb_{21}} = a^{4y}$ (resp., $u_3^{yb_{23}} = a^y$) is always 1 (resp., 3). Observe that $\varphi(B) \subseteq a^*\mathbf{c}a^*\mathbf{c}a^*u$, where $u = \mathbf{c}a^7\mathbf{c}a^2$. Let $\widehat{B}$ be the simple set of $\mathbb{N}^5$ given by $\widehat{B} = \widehat{\mathbf{b}}_0 + \{\widehat{\mathbf{b}}_1, \widehat{\mathbf{b}}_2\}^\oplus$, with $\widehat{\mathbf{b}}_0 = (1, 3, 2, 7, 2)$, $\widehat{\mathbf{b}}_1 = (4, 0, 0, 0, 0)$, and $\widehat{\mathbf{b}}_2 = (4, 0, 1, 0, 0)$. The reader can check that $\widehat{\varphi}(\widehat{B}) = \varphi(B)$ with respect to the map $\widehat{\varphi} : \mathbb{N}^5 \longrightarrow a^*\mathbf{c}a^*\mathbf{c}a^*\mathbf{c}a^*\mathbf{c}a^*$. Finally observe that $\widehat{B}$ can be written as $\widehat{B} = D \times 0^2 + \mathbf{b}_u$, where $\mathbf{b}_u = (0, 0, 0, 7, 2)$ and D is the simple set of $\mathbb{N}^3$ given by $D = \widehat{\widehat{\mathbf{b}}}_0 + \{\widehat{\widehat{\mathbf{b}}}_1, \widehat{\widehat{\mathbf{b}}}_2\}^\oplus$, where $\widehat{\widehat{\mathbf{b}}}_0 = (1, 3, 2)$, $\widehat{\widehat{\mathbf{b}}}_1 = (4, 0, 0)$, $\widehat{\widehat{\mathbf{b}}}_2 = (4, 0, 1)$.