Theoretical Computer Science 562 (2015) 304–319

Contents lists available at ScienceDirect

Theoretical Computer Science

www.elsevier.com/locate/tcs

On the commutative equivalence of bounded context-free and regular languages: The code case

Flavio D'Alessandro [a,b,*], Benedetto Intrigila [c]

[a] Dipartimento di Matematica, Università di Roma "La Sapienza", Piazzale Aldo Moro 2, 00185 Roma, Italy
[b] Department of Mathematics, Boğaziçi University, 34342 Bebek, Istanbul, Turkey
[c] Dipartimento di Ingegneria dell'Impresa, Università di Roma "Tor Vergata", via del Politecnico, 1, 00133 Roma, Italy

ARTICLE INFO

Article history:
Received 20 January 2014
Received in revised form 23 September 2014
Accepted 1 October 2014
Available online 8 October 2014
Communicated by G. Ausiello

Keywords:
Bounded context-free language
Semilinear set
Commutative equivalence

ABSTRACT

This is the first 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.

1. Introduction

In this paper, we study an algebraic and combinatorial problem concerning bounded context-free languages. A strictly related notion is that of *sparse language*: a language L is termed *sparse* if its counting function, that is, the function f_L that maps every integer $n \geq 0$ into the number $f_L(n)$ of words of L of length n, is polynomially upper bounded. Sparse languages play a meaningful role both in Computer Science and in Mathematics and have been widely investigated in the past. The interest in this class of languages is due to the fact that, in the context-free case, it coincides with the one of *bounded languages*. A language L is termed *bounded* if there exist k words $u_1, \ldots, u_k$ such that $L \subseteq u_1^* \cdots u_k^*$ ([3,5,6,14,16–20,22,23,25]; see also [11] for an excellent survey on the relationships between bounded languages and monoids of polynomial growth).

The starting point of this investigation is the following result proved in [5]: for every sparse context-free language L_1, there exists a regular language L_2 such that $f_{L_1} = f_{L_2}$.

Therefore, the counting function of a sparse context-free language is always rational. This result is interesting since, as it is well known [13], the counting function of a context-free language may be transcendental.

A conceptual limit of the above-mentioned construction is that the regular language L_2 is on a different alphabet from the one of L_1. Therefore it is natural to ask whether this limitation can be removed. It is worth noticing that, from a general point of view, imposing restrictions on the alphabets makes some classical constructions more difficult but obviously more informative. In this context, as a related result, we can mention a remarkable contribution by Beal and Perrin where the problem of the length equivalence of regular languages on alphabets of prescribed size is considered [1].

Let us now describe the problem we are interested in. For this purpose, some preliminary notions are needed. 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

* Corresponding author.

http://dx.doi.org/10.1016/j.tcs.2014.10.005
0304-3975/

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 notion is important in the theory of variable-length codes since it is involved in the celebrated Schützenberger conjecture about the commutative equivalence of a maximal finite code with a prefix one (see e.g. [2]).

Now the general problem above can be formulated as follows: *given a bounded context-free language L_1, does it exist a regular language L_2 which is commutatively equivalent to L_1?*
In the sequel, for short, we refer to it as *CE (Commutative Equivalence) Problem*.

This is the first paper of a group of three (see also [9,10]) where we prove the following statement that solves in the affirmative the CE Problem.

Theorem 1. *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.*

Theorem 1 with a sketch of the proof was announced in [7]. Actually we prove that the CE problem can be solved in the affirmative for the wider class of bounded semi-linear languages. Moreover the use of such languages turns out to be of crucial importance in the solution of the problem as it makes possible to handle the languages through suitable sets of vectors of integers, namely semi-linear sets of the free commutative monoid $\mathbb{N}^k$.

Observe that the CE Problem can be seen as a kind of counting problem in the class of bounded context-free languages. Despite the fact that such class has been widely investigated in the past, CE Problem appears to require some new techniques. Indeed, bounded context-free languages can be inherently ambiguous; in addition, if $u_1^* \cdots u_k^*$, $u_i \in A^+$, is the set that contains the bounded context-free language, then, in general, $u_1^* \cdots u_k^*$ is ambiguous as product of languages of A^*. Such ambiguities, which are of different nature, interfere making the study of the problem a non-trivial task.

Before explaining the main contribution of this paper, we would like to give a broader picture about the relationships between CE Problem and some classical theorems on bounded context-free languages. The first result that deserves to be mentioned is a well-known theorem by Parikh [24]. For this purpose, let us first introduce a notion. Given two languages L_1 and L_2 over the alphabet A, we say that L_1 is *Parikh equivalent to* L_2 if $\psi(L_1) = \psi(L_2)$. The theorem by Parikh states that, given a context-free language L_1, its image $\psi(L_1)$ under the Parikh map is a semi-linear set of $\mathbb{N}^t$. As a straightforward consequence of Parikh theorem, one has that there exists a regular language L_2 which is *Parikh equivalent to* L_1.

It is worth noticing that the property of Parikh equivalence by no means implies the property of commutative equivalence. Indeed, let $A = \{a, b\}$ and let $L_1 = (ab)^* \cup (ba)^*$ and $L_2 = (ab)^*$. One has that $\psi(L_1) = \psi(L_2) = (1, 1)^{\oplus}$ (the symbol $\oplus$ denotes the Kleene star operation in the monoid $\mathbb{N}^2$) so that L_1 is Parikh equivalent to L_2. On the other hand, one immediately checks that L_1 cannot be commutatively equivalent to L_2.

Another theorem that is central in this context has been proved by Ginsburg and Spanier [15]. For this purpose, let us first introduce a notion. Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded language where, for every $i = 1, \ldots, k$, u_i is a word over the alphabet A. Let $\varphi : \mathbb{N}^k \longrightarrow u_1^* \cdots u_k^*$ be the map defined as: for every tuple $(\ell_1, \ldots, \ell_k) \in \mathbb{N}^k$,

$$\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}.$$

The map φ is called the *Ginsburg map*. Ginsburg and Spanier proved that L is context-free if and only if $\varphi^{-1}(L)$ is a finite union of linear sets, each having a stratified set of periods. Roughly speaking, a stratified set of periods corresponds to a system of well-formed parentheses. However, Ginsburg and Spanier theorem is of no help to study counting problems and, in particular, our problem, because of the ambiguity of the representation of such languages. Indeed, let $A = \{a, b, c\}$ be a three letter-alphabet and let the language $L = \{a^i b^j c^k \mid i, j, k \in \mathbb{N},\ i = j \text{ or } j = k\}$ [4]. Since L is inherently ambiguous, by [14] Theorem 6.2.1, L cannot be represented unambiguously as a finite disjoint union of a stratified set of periods. In this context, another important recent result that gives a characterization of bounded context-free languages in terms of finite unions of *Dyck loops* has been proven in [20]. However, neither this latter result can be used to deal with our problem because of the ambiguity of the representation of such languages as a finite union of Dyck loops.

The proof of Theorem 1 will be presented in its complete generality in [10]. In particular, we will prove Theorem 1, by using a refined version of the techniques of this paper together with another result proved in [9].

The goal of this paper is to describe some key elements of our technique. We will do this by proving Theorem 1 under the following assumption.

Let $L = \varphi(B)$ be a language described, via the Ginsburg map, by a semi-simple set B:

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

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

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

where $k_i > 0$ is the dimension of B_i and the vectors $\mathbf{b}_0^{(i)}, \mathbf{b}_1^{(i)}, \ldots, \mathbf{b}_{k_i}^{(i)}$, form the unambiguous representation of B_i.

The main contribution of this paper can be stated as follows.

Theorem 2. *Suppose that, for every $i = 1, \dots, n$ and for every $j = 1, \dots, k_i$, the word $\varphi(\mathbf{b}_j^{(i)})$ has at least two distinct letters in its factorization.*

Then L is commutatively equivalent to a regular language L'. Moreover the language L' can be effectively constructed starting from an effective presentation of L.

Two main tools are used in the proof of Theorem 2. The first tool is a combinatorial argument based on uniform codes that makes possible to separate the languages that represent the simple sets B_i of the decomposition of B.

The second one is a technique (inspired by our work [8]) of geometrical nature for the decomposition into parallelepipeds of simple sets.

We first prove Theorem 2 in the case that the language L is described by a simple set (Section 4.1); afterwards we extend such solution to the case of a language L described by a semi-simple set (Section 4.2).

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

2. Preliminaries and basic definitions

The aim of this section is to introduce some results concerning bounded context-free and semi-linear languages. We assume that the reader is familiar with the basic notions of context-free languages. The reader is referred to [14].

2.1. Basic notation

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

The letter k will be 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 $(v_1, \dots, v_k)$. Moreover if the vector is indexed, as for instance in $\mathbf{v}_j$, its components are denoted $(v_{j1}, v_{j2}, \dots, v_{jk})$, or $\mathbf{v}_j = (v_1^{(j)}, v_2^{(j)}, \dots, v_k^{(j)})$.

A subset of $\mathbb{N}^k$ or of A^* is always denoted by using capital letters like, for instance, X, Y, L, *etc.* We use Greek letters, as for instance, α, β, γ to denote constants.

An (indexed) family of sets is denoted by a calligraphic letter like, for instance, $\mathcal{X} = \{X_i\}_{i \ge 1}$, and $\mathcal{Y} = \{Y_i\}_{i \ge 1}$.

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

We start by introducing some basic results concerning 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 $B = \{\mathbf{b}_1, \dots, \mathbf{b}_m\}$ 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

$$B^\oplus = \mathbf{b}_1^\oplus + \cdots + \mathbf{b}_m^\oplus = \{n_1\mathbf{b}_1 + \cdots + n_m\mathbf{b}_m \mid n_i \in \mathbb{N}\}.$$

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 $X = \mathbf{b}_0 + \{\mathbf{b}_1, \dots, \mathbf{b}_m\}^\oplus$, where $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_m$ are vectors of $\mathbb{N}^k$,
2. X is *simple* in $\mathbb{N}^k$ if the vectors $\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_m$ 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 $\mathbb{N}^k$.

In the sequel, we will adopt the following terminology.

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

Indeed, according to *folklore*, one can prove that the representation of a simple set is unique. With every simple set B whose representation is given by the vectors $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_m$, we can associate the number m called the *dimension* of B. One obviously has that $m \le k$ 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 [12] provides an important characterization of semi-linear sets. Also it is worth to remark that Theorem 3 was proved independently by Ito in [21].

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 the representation of a simple set X_i and X is the disjoint union of the sets X_i.

The following proposition states a basic result for semi-linear sets.

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

2.3. Bounded languages

In 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 denote the family of regular languages and the family of context-free languages by $Rat(A^*)$ and $CFL(A^*)$ respectively. We let

$$\psi : A^* \longrightarrow \mathbb{N}^t, \tag{1}$$

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 map defined as: for every tuple $(\ell_1, \ldots, \ell_k) \in \mathbb{N}^k$,

$$\varphi(\ell_1, \ldots, \ell_k) = u_1^{\ell_1} \cdots u_k^{\ell_k}.$$

The map φ is called the *Ginsburg map*.

Let us consider the map:

$$\varphi\psi : \mathbb{N}^k \longrightarrow \mathbb{N}^t,$$

such that, for every $\mathbf{b} \in \mathbb{N}^k$, $(\varphi\psi)(\mathbf{b}) = \psi(\varphi(\mathbf{b}))$. The following lemma is immediate.

Lemma 1. *For every $\mathbf{v}, \mathbf{v}' \in \mathbb{N}^k$, one has $(\varphi\psi)(\mathbf{v} + \mathbf{v}') = \psi(\varphi(\mathbf{v})\varphi(\mathbf{v}'))$.*

The following proposition, proved in [17], provides a remarkable tool to describe, in terms of semi-linear sets, a bounded context-free language (see also [5,6] for a proof).

Proposition 2. *Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded context-free 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

We display some notions and elementary results of Combinatorics on words that are needed in this setting. We now recall the definition of commutative equivalence of two languages.

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 lemma is an easy consequence of the previous definition.

Lemma 2. *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 $(L_1 \cup L_2) \sim (L_1' \cup L_2')$.*

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 in the rest of the paper.

The following lemma is very useful in our setting and its proof is easy (for the proof see Appendix A).

Lemma 3. *(Non-factorization lemma) 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*

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

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

$$z_1, \ldots, z_m, \tag{4}$$

be a list of (not necessarily pairwise distinct) words of A^+. We associate with it its corresponding multiset of Parikh vectors:

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

where $\alpha_1 + \cdots + \alpha_\ell = m$ and, for every $i = 1, \ldots, \ell$, α_i is the number of words of the list whose Parikh vector is $\mathbf{v}_i$.

We recall that a code over A^* is said to be *uniform* if all the words of the set have the same length. The following lemma holds.

Lemma 4. *Let us consider the list of words (4) together with its multiset of Parikh vectors (5). Suppose that:*

(i) *for every $j = 1, \ldots, m$, z_j contains, at least, two different letters of A in its factorization;*
(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

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

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 word $w \in \mathcal{W}$ is not a factor of any word in $u_1^* \cdots u_k^*$.

Proof. Recall that $A = \{a_1, \ldots, a_t\}$. We consider first the case where for every $i = 1, \ldots, \ell$, $\alpha_i = 1$. Hence $m = \ell$. Given the Parikh vector $\mathbf{v}_i = (v_1^{(i)}, v_2^{(i)}, \ldots, v_t^{(i)})$, with $1 \leq i \leq \ell$, by (i) there exist two components, say $v_1^{(i)}$ and $v_2^{(i)}$ for simplicity, which are different from 0. Since by hypothesis, $v_j^{(i)} \geq m(\gamma + 1)(k + 1)$, for $j = 1, 2$, we can consider the word

$$t_{\mathbf{v}_i} = a_1^\gamma a_2 a_1^\gamma a_2 \ldots a_1^\gamma a_2 a_1^\gamma$$

with a_1^γ occurring $(k+1)$-times in $t_{\mathbf{v}_i}$. With the vector $\mathbf{v}_i$ we associate the word of A^+:

$$w_{\mathbf{v}_i} = t_{\mathbf{v}_i} a_1^{(v_1^{(i)} - \gamma(k+1))} a_2^{(v_2^{(i)} - k)} a_3^{v_3^{(i)}} \cdots a_t^{v_t^{(i)}}.$$

Repeating this construction for every vector $\mathbf{v}_i$, $i = 1, \ldots, \ell$, and observing that in our case $m = \ell$, we obtain m words $w_{\mathbf{v}_i}$. Let $\mathcal{W}$ be the set of such words. By construction $\mathcal{W}$ satisfies (6) and, by (ii), all the words of $\mathcal{W}$ have the same length β. Moreover, since $\alpha_i = 1$, $i = 1, \ldots, \ell$, all the words of $\mathcal{W}$ are obviously distinct. Thus $\mathcal{W}$ is a uniform code of m words. Finally, by Lemma 3, for every $w_{\mathbf{v}_i}$, $w_{\mathbf{v}_i}$ cannot be a factor of a word in $u_1^* \cdots u_k^*$, since $t_{\mathbf{v}_i}$ is a factor of $w_{\mathbf{v}_i}$.

Now let us treat the case where for some i with $i = 1, \ldots, \ell$, $\alpha_i > 1$. Let $\mathbf{v}_i = (v_1^{(i)}, v_2^{(i)}, \ldots v_t^{(i)})$, $i = 1, \ldots, \ell$. By (i) there exist two components, say $v_1^{(i)}$ and $v_2^{(i)}$ for simplicity, which are different from 0. Since, moreover, $v_j^{(i)} \geq m(\gamma + 1)(k + 1)$, for $j = 1, 2$, we can consider the word

$$t_{\mathbf{v}_i} = a_1^\gamma a_2 a_1^\gamma a_2 \ldots a_1^\gamma a_2 a_1^\gamma$$

with a_1^γ occurring $(k+1)$-times in $t_{\mathbf{v}_i}$. Now since $v_2^{(i)} \geq m(\gamma + 1)(k + 1)$, one has $(v_2^{(i)} - k) > m \geq \alpha_i$, so that we can define the following α_i words:

$$t_{\mathbf{v}_i} a_1^{(v_1^{(i)} - \gamma(k+1))} a_2^{(v_2^{(i)} - k)} \cdots a_p^{v_p^{(i)}} \cdots a_t^{v_t^{(i)}},$$

$$t_{\mathbf{v}_i} a_2 a_1^{(v_1^{(i)} - \gamma(k+1))} a_2^{(v_2^{(i)} - k - 1)} \cdots a_p^{v_p^{(i)}} \cdots a_t^{v_t^{(i)}},$$

$$t_{\mathbf{v}_i} a_2^2 a_1^{(v_1^{(i)} - \gamma(k+1))} a_2^{(v_2^{(i)} - k - 2)} \cdots a_p^{v_p^{(i)}} \cdots a_t^{v_t^{(i)}},$$

.

.

$$t_{\mathbf{v}_i} a_2^{\alpha_i - 1} a_1^{(v_1^{(i)} - \gamma(k+1))} a_2^{(v_2^{(i)} - k - \alpha_i + 1)} \cdots a_p^{v_p^{(i)}} \cdots a_t^{v_t^{(i)}}.$$

Now we associate with $\mathbf{v}_i$ the set $\mathcal{W}_{\mathbf{v}_i}$ of such α_i words. One immediately verifies that, by construction of the set $\mathcal{W}_{\mathbf{v}_i}$, all the words of the set are pairwise distinct and moreover, for every $w \in \mathcal{W}_{\mathbf{v}_i}$, $\psi(w) = \mathbf{v}_i$ and thus $|w| = \beta$.

Finally, repeating this construction for every vector $\mathbf{v}_i$, for $i = 1, ..., \ell$, we get a set of words $\mathcal{W}$ of A^+ defined as $\mathcal{W} = \bigcup_{i=1}^{\ell} \mathcal{W}_{\mathbf{v}_i}$. It is easily seen that all the words of $\mathcal{W}$ are pairwise distinct. Indeed, let $w', w \in \mathcal{W}$. Then either $w', w \in \mathcal{W}_{\mathbf{v}_i}$, with $i = 1, \ldots, \ell$ or $w' \in \mathcal{W}_{\mathbf{v}_i}, w \in \mathcal{W}_{\mathbf{v}_j}$, with $1 \le i < j \le \ell$. In the first case, w' and w are distinct because of the construction of the set $\mathcal{W}_{\mathbf{v}_i}$. In the second case, one has $\psi(w') = \mathbf{v}_i \neq \mathbf{v}_j = \psi(w)$ and therefore $w' \neq w$. As an immediate consequence of the previous argument, one has that $\mathcal{W}$ is a uniform code of m (distinct) words of A^+ of length β satisfying (6). By Lemma 3, for every $w \in \mathcal{W}$, w cannot be a factor of a word in $u_1^* \cdots u_k^*$. Indeed, if $w \in \mathcal{W}_{\mathbf{v}_i}$, for some i, then it follows that $t_{\mathbf{v}_i}$ is a factor of w. This completes the proof. □

Example 1. This example shows an application of the previous lemma (see Appendix A).

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 have two different symbols in their factorization, at least. Under the assumption that all the components of every Parikh vector are sufficiently large, then 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 a given context-free language contained in $u_1^* \cdots u_k^*$ (cf. Lemma 8).

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

In this section, we introduce a new technique (inspired by our work [8]) of geometrical nature for the decomposition of simple sets. Let B be a simple set of $\mathbb{N}^k$ of dimension $m > 0$:

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

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

Let

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

be a sequence of m positive integers.

From an intuitive point of view, the goal of this section is to define, according to a sequence of m positive integers, a suitable decomposition of B into parallelepipeds of dimension lower than or equal to m.

Let $\{+, -\}$ be an alphabet of two symbols and let $\mathcal{E}$ be the set

$$\mathcal{E} = \left\{(\epsilon_1, \ldots, \epsilon_m) \;\middle|\; \epsilon_i \in \{+, -\},\ i = 1, \ldots, m\right\},$$

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 of vectors $B^{(\epsilon_1, \ldots, \epsilon_m)}$ defined as:

$$B^{(\epsilon_1, \ldots, \epsilon_m)} = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m \mid x_i \in \mathbb{N}\}, \tag{8}$$

where, for every $i = 1, \ldots, m$, one has:

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

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^-.

Lemma 5. *The following conditions hold:*

1. *For every $(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E}$, $B^{(\epsilon_1, \ldots, \epsilon_m)}$ is a semi-simple set;*
2. *B^- is the unique finite set of the family (8);*

3. *The family of vectors defined in (8):*

$$\left\{B^{(\epsilon_1,\dots,\epsilon_m)}\right\}_{(\epsilon_1,\dots,\epsilon_m)\in\mathcal{E}}$$

gives a partition:

$$B = \bigcup_{(\epsilon_1,\dots,\epsilon_m)\in\mathcal{E}} B^{(\epsilon_1,\dots,\epsilon_m)}$$

of B.

Proof. All the claims follow immediately from definition (8) and from the fact that the vectors $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_m$ form the representation of the simple set B. $\square$

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

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

where

$$\epsilon_{i_\ell} = -, \quad (\ell = 1, \dots, p), \qquad \epsilon_{i_\ell} = +, \quad (\ell = p+1, \dots, m).$$

It's worthy to remark that:

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

Let us now associate with every index i_ℓ of the set $I^+_{\epsilon_1\cdots\epsilon_m} = \{i_{p+1}, \dots, i_m\}$ of (9) a remainder r_{i_ℓ} modulo N_{i_ℓ}, where all the constants N_{i_ℓ} are defined in (7). We thus obtain a sequence of $m - p$ constants

$$(r_{i_{p+1}}, \dots, r_{i_m}). \tag{10}$$

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

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

$$\mathbf{b}_0 + \left\{\sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} \,\middle|\, 0 \le c_{i_\ell} < N_{i_\ell}\right\} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell} + \left\{\sum_{\ell=p+1}^{m} N_{i_\ell}x_{i_\ell}\mathbf{b}_{i_\ell} \,\middle|\, x_{i_\ell} \ge 1\right\}. \tag{11}$$

Observe that $B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}$ is the finite union of the pairwise disjoint sets

$$\left\{\mathbf{b}_0 + \sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} N_{i_\ell}x_{i_\ell}\mathbf{b}_{i_\ell} \,\middle|\, x_{i_\ell} \ge 1\right\},$$

where $(c_{i_1}, \dots, c_{i_p})$ is a tuple of integers where, for every $\ell = 1, \dots, p$, $0 \le c_{i_\ell} < N_{i_\ell}$.

Lemma 6. *Let $(\epsilon_1, \dots, \epsilon_m) \in \mathcal{E} \setminus \{(-, -, \dots, -)\}$. The family of vectors $B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}$, with $(r_{i_{p+1}}, \dots, r_{i_m}) \in \mathcal{C}^+_{\epsilon_1\cdots\epsilon_m}$, defined in (11), is a partition of the set $B^{(\epsilon_1,\dots,\epsilon_m)}$ into semi-simple sets*

$$B^{(\epsilon_1,\dots,\epsilon_m)} = \bigcup_{(r_{i_{p+1}},\dots,r_{i_m})\in\mathcal{C}^+_{\epsilon_1\cdots\epsilon_m}} B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}.$$

Proof. All the claims follow immediately from definition (11) and from the fact that the vectors $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_m$ form the representation of the simple set B. $\square$

As an immediate consequence of the previous two lemmas, we obtain the following proposition.

Proposition 3. *The simple set $B = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m \mid x_i \in \mathbb{N}\}$ admits the partition into semi-simple sets*

$$B^- \cup \bigcup_{(\epsilon_1,\dots,\epsilon_m)\in\mathcal{E}\setminus\{(-,-,\dots,-)\}} \; \bigcup_{(r_{i_{p+1}},\dots,r_{i_m})\in\mathcal{C}^+_{\epsilon_1\cdots\epsilon_m}} B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}, \tag{12}$$

constructed starting from the sequence (7) of numbers N_i, $i = 1, \dots, m$.

Example 2. This example shows an application of Proposition 3 (see Appendix A).

Let

$$L^- = \varphi\big(B^-\big),$$

and let us define the following family of languages of A^*:

$$L^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}} = \varphi\big(B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}\big), \tag{13}$$

for every $(\epsilon_1,\dots,\epsilon_m) \in \mathcal{E} \setminus \{(-,-,\dots,-)\}$ and for every $(r_{i_{p+1}},\dots,r_{i_m}) \in \mathcal{C}^+_{\epsilon_1\cdots\epsilon_m}$.

Then the following proposition holds.

Proposition 4. *The family of languages (13), together with L^-, gives a partition of L.*

Proof. The claim immediately follows from Proposition 3 and Proposition 2. □

4. The construction of the regular language

In this section, we prove Theorem 2. Let $L \subseteq u_1^* \cdots u_k^*$ be a bounded context-free 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-linear set B of $\mathbb{N}^k$ such that $\varphi(B) = L$ and φ is injective on B. Moreover, by Theorem 3, we can assume that B is semi-simple.

We start to prove Theorem 2 under the assumption that the language L is described, via the map φ, by a simple set. This will be done in the next section.

4.1. The simple case

In the sequel of this section we will do the following assumption:

Assumption 1.

- there exists a simple set B of $\mathbb{N}^k$ of dimension $m > 0$ such that $\varphi(B) = L$ and φ is injective on B.
- if B is written as:

  $$B = \{\mathbf{b}_0 + x_1\mathbf{b}_1 + \cdots + x_m\mathbf{b}_m \mid x_i \in \mathbb{N}\},$$

  where the vectors $\mathbf{b}_0, \mathbf{b}_1, \dots, \mathbf{b}_m$, form the representation of B, then, for every $i = 1,\dots,m$, the word $\varphi(\mathbf{b}_i)$ has, at least, two distinct letters in its factorization.

Remark 2. It is useful to remark that, since for every $i = 1,\dots,m$, $\mathbf{b}_i$ is not the null vector, $\varphi(\mathbf{b}_i) \neq 1_{A^*}$. Moreover, the words $\varphi(\mathbf{b}_i)$, $i = 1,\dots,m$, may be not all pairwise distinct. Therefore, in the sequel, the words

$$\varphi(\mathbf{b}_1),\dots,\varphi(\mathbf{b}_m) \tag{14}$$

will be conceived as a list: two possibly equal words have to be considered distinct if they have different positions into the sequence.

Let c be a non-negative integer and let $\beta(c)$ be defined as

$$\beta(c) = \prod_{\ell=1}^{m} \big|\varphi(\mathbf{b}_\ell)\big| c.$$

For every $i = 1,\dots,m$, let $N_i(c)$ be the number defined as:

$$N_i(c) = \left(\frac{\beta(c)}{|\varphi(\mathbf{b}_i)|}\right) = \left(\frac{\prod_{\ell=1}^{m} |\varphi(\mathbf{b}_\ell)| c}{|\varphi(\mathbf{b}_i)|}\right). \tag{15}$$

Lemma 7. *Let $N_i(c)$, $i = 1,\dots,m$, be the numbers defined in (15). For every $i = 1,\dots,m$, one has:*

$$\big|\varphi\big(N_i(c)\mathbf{b}_i\big)\big| = \beta(c).$$

Proof. Remark first that, for every $x \in \mathbb{N}$ and for every $\mathbf{b} \in \mathbb{N}^k$, $|\varphi(x\mathbf{b})| = x|\varphi(\mathbf{b})|$. Hence, by (15), for every $i = 1, \ldots, m$, one has:

$$\left|\varphi\left(N_i(c)\mathbf{b}_i\right)\right| = N_i(c)\left|\varphi(\mathbf{b}_i)\right| = \left(\frac{\prod_{\ell=1}^{m}|\varphi(\mathbf{b}_\ell)|}{|\varphi(\mathbf{b}_i)|}\right)c\left|\varphi(\mathbf{b}_i)\right| = \beta(c),$$

and the claim is thus achieved. □

From now on, we will assume that c is a positive integer such that, for every $i = 1, \ldots, k$:

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

where m is the number of the vectors of the representation of B and γ is the fixed constant of Lemma 3. The list of (possibly equal) words

$$\varphi\left(N_1(c)\mathbf{b}_1\right), \ldots, \varphi\left(N_m(c)\mathbf{b}_m\right) \tag{17}$$

satisfy the hypotheses of Lemma 4. Hence by applying this lemma to the previous list of words (17), one gets the following lemma.

Lemma 8. *There exists a uniform length code $\mathcal{W}$ of m distinct words of length $\beta(c)$:*

$$\mathcal{W} = \{w_1, \ldots, w_m\},$$

such that, for every $i = 1, \ldots, m$,

$$\psi(w_i) = \psi\left(\varphi\left(N_i(c)\mathbf{b}_i\right)\right).$$

Moreover, for every $i = 1, \ldots, m$, w_i is not a factor of any word of $u_1^* \cdots u_k^*$.

By Lemma 8, there exists a uniform length code of m distinct words

$$\mathcal{W} = \{w_1, \ldots, w_m\}, \tag{18}$$

where, for every $i = 1, \ldots, m$,

$$\psi(w_i) = \psi\left(\varphi\left(N_i(c)\mathbf{b}_i\right)\right). \tag{19}$$

Let us consider the bijective map

$$N_i(c)\mathbf{b}_i \longrightarrow w_i, \tag{20}$$

that maps the i-th vector $N_i(c)\mathbf{b}_i$ into the corresponding i-th word of $\mathcal{W}$, for every $i = 1, \ldots, m$.

Remark 3. Let $w_i \in \mathcal{W}$, $i = 1, \ldots, m$, be a word of the code (18).

(i) For every $i = 1, \ldots, m$, w_i 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 word $w \in \mathcal{W}$ is not a factor of any word in $u_1^* \cdots u_k^*$.
(ii) By (19), $|w_i| = |\varphi(N_i(c)\mathbf{b}_i)| = \beta(c)$.

For the sake of simplicity, from now on, the numbers $N_1(c), \ldots, N_m(c)$ and $\beta(c)$ will be denoted respectively as:

$$N_1, \ldots, N_m, \ \beta. \tag{21}$$

Let us now consider the partition (12) of B, given by Proposition 3, with respect to the sequence

$$(N_1, \ldots, N_m). \tag{22}$$

In order to simplify the notation, let us fix an enumeration:

$$\mathcal{B}_1, \ldots, \mathcal{B}_s, \quad s \geq 1,$$

of the sets of (12) distinct from B^-. Hence, for every $(\epsilon_1, \ldots, \epsilon_m) \in \mathcal{E} \setminus (-, -, \cdots, -)$, and for every $(r_{i_{p+1}}, \ldots, r_{i_{m-p}}) \in \mathcal{C}^+_{\epsilon_1 \cdots \epsilon_m}$, there exists exactly one index i, with $1 \leq i \leq s$, such that $\mathcal{B}_i = B^{(\epsilon_1, \ldots, \epsilon_m)}_{r_{i_{p+1}} \cdots r_{i_m}}$ is the set:

$$\mathbf{b}_0 + \left\{\sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} \,\middle|\, 0 \leq c_{i_\ell} < N_{i_\ell}\right\} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell} + \left\{\sum_{\ell=p+1}^{m} N_{i_\ell}x_{i_\ell}\mathbf{b}_{i_\ell} \,\middle|\, x_{i_\ell} \geq 1\right\}.$$

Let us associate with every set B_i, with $i = 1, \ldots, s$, the following regular language:

$$L'_i = \varphi\left(\mathbf{b}_0 + \left\{\sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} \;\middle|\; 0 \le c_{i_\ell} < N_{i_\ell}\right\} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell}\right) w^+_{i_{p+1}} w^+_{i_{p+2}} \cdots w^+_{i_m}, \tag{23}$$

where, for every $\ell = p+1, \dots, m$, w_{i_ℓ} is the word of the set (18) associated with the coding (20) to the index i_ℓ.

Observe that L'_i is the finite union of the regular languages

$$\varphi\left(\mathbf{b}_0 + \sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell}\right) w^+_{i_{p+1}} w^+_{i_{p+2}} \cdots w^+_{i_m},$$

where the sequence of code words $w_{i_{p+1}}, w_{i_{p+2}}, \dots, w_{i_m}$, is fixed and $(c_{i_1}, \dots, c_{i_p})$ is a tuple of integers where, for every $\ell = 1, \dots, p$, $0 \le c_{i_\ell} < N_{i_\ell}$.

In order to keep our notation uniform, we denote the languages of the family (13) as:

$$L_i = \varphi(\mathcal{B}_i), \tag{24}$$

for every $i = 1, \dots, s$.

Lemma 9. *The languages of the family (23) are pairwise disjoint. Moreover every language of the family (23) is disjoint with $L^- = \varphi(B^-)$.*

Proof. Let us prove the first part of the claim. By contradiction, assume that there exist two languages L'_i and L'_j of the family (23), with $1 \le i < j \le s$, such that $L'_i \cap L'_j \ne \emptyset$. Then the language L'_i is defined as:

$$L'_i = \varphi\left(\mathbf{b}_0 + \left\{\sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} \;\middle|\; 0 \le c_{i_\ell} < N_{i_\ell}\right\} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell}\right) w^+_{i_{p+1}} w^+_{i_{p+2}} \cdots w^+_{i_m}, \tag{25}$$

and it is associated with the set of vectors:

$$\mathcal{B}_i = B^{(\epsilon_1,\dots,\epsilon_m)}_{r_{i_{p+1}}\cdots r_{i_m}}.$$

Similarly the language L'_j is defined as:

$$L'_j = \varphi\left(\mathbf{b}_0 + \left\{\sum_{\ell=1}^{q} d_{j_\ell}\mathbf{b}_{j_\ell} \;\middle|\; 0 \le d_{j_\ell} < N_{j_\ell}\right\} + \sum_{\ell=q+1}^{m} s_{j_\ell}\mathbf{b}_{j_\ell}\right) w^+_{j_{q+1}} w^+_{j_{q+2}} \cdots w^+_{j_m}, \tag{26}$$

and it is associated with the set of vectors:

$$\mathcal{B}_j = B^{(\delta_1,\dots,\delta_m)}_{s_{j_{q+1}}\cdots s_{j_m}}.$$

By hypothesis, there exists a word $u \in L'_i \cap L'_j \ne \emptyset$. Now we prove that $\mathcal{B}_i = \mathcal{B}_j$, which is a contradiction.

Since $u \in L'_i$, by (25), there exist exactly two tuples

$$x_{i_{p+1}}, x_{i_{p+2}}, \dots, x_{i_m} \ge 1$$

and

$$c_{i_1}, \dots, c_{i_p},$$

with, for every $\ell = 1, \dots, p$, $0 \le c_{i_\ell} < N_{i_\ell}$ such that:

$$u = \varphi(\mathbf{v}_i) w^{x_{i_{p+1}}}_{i_{p+1}} w^{x_{i_{p+2}}}_{i_{p+2}} \cdots w^{x_{i_m}}_{i_m},$$

where

$$\mathbf{v}_i = \mathbf{b}_0 + \sum_{\ell=1}^{p} c_{i_\ell}\mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} r_{i_\ell}\mathbf{b}_{i_\ell}.$$

Similarly, since $u \in L'_j$, by (26), there exist exactly two tuples

$$y_{i_{q+1}}, y_{i_{q+2}}, \dots, y_{i_m} \ge 1$$

and

$$d_{i_1}, \dots, d_{i_q},$$

with, for every $\ell = 1, \dots, q$, $0 \le d_{j_\ell} < N_{j_\ell}$ such that

$$u = \varphi(\mathbf{v}_j) w_{j_{q+1}}^{y_{j_{q+1}}} w_{j_{q+2}}^{y_{j_{q+2}}} \cdots w_{j_m}^{y_{j_m}},$$

where

$$\mathbf{v}_j = \mathbf{b}_0 + \sum_{\ell=1}^{q} d_{j_\ell} \mathbf{b}_{j_\ell} + \sum_{\ell=q+1}^{m} s_{j_\ell} \mathbf{b}_{j_\ell}.$$

Set

$$U_1 = w_{i_{p+1}}^{x_{i_{p+1}}} w_{i_{p+2}}^{x_{i_{p+2}}} \cdots w_{i_m}^{x_{i_m}}, \qquad U_2 = w_{j_{q+1}}^{y_{j_{q+1}}} w_{j_{q+2}}^{y_{j_{q+2}}} \cdots w_{j_m}^{y_{j_m}}.$$

Let us first show that $|U_1| = |U_2|$. Indeed, assume $|U_1| < |U_2|$ (the other case is treated similarly). Since $U_1, U_2 \in \mathcal{W}^+$ and all the words of $\mathcal{W}$ have the same length, the latter implies that there exists a word $w \in \mathcal{W}$ such that w is a suffix of $\varphi(\mathbf{v}_i)$. Since $\varphi(\mathbf{v}_i) \in u_1^* \cdots u_k^*$ and since (cf. Remark 3) w cannot be a factor of any word of $u_1^* \cdots u_k^*$, one gets a contradiction. Hence $|U_1| = |U_2|$. The latter implies that

$$U_1 = U_2, \qquad \varphi(\mathbf{v}_i) = \varphi(\mathbf{v}_j).$$

From $\varphi(\mathbf{v}_i) = \varphi(\mathbf{v}_j)$ one has $\mathbf{v}_i = \mathbf{v}_j$. Indeed, $\mathbf{v}_i, \mathbf{v}_j \in B$ and φ is injective on B.

From $U_1 = U_2 \in \mathcal{W}^+$, since $\mathcal{W}$ is a code, one has that the two factorizations U_1 and U_2 must be equal. This implies that

$$p = q, \qquad i_{p+1} = j_{q+1}, \qquad i_{p+2} = j_{q+2}, \ldots, i_m = j_m,$$

and

$$x_{i_{p+1}} = y_{j_{q+1}}, \qquad x_{i_{p+2}} = y_{j_{q+2}}, \ldots, x_{i_m} = y_{j_m}.$$

Now recall that, by the bijective map (20), every word of $\mathcal{W}$ codifies exactly one vector $N_\ell \mathbf{b}_\ell$, $1 \le \ell \le m$. Hence, from the latter condition, we get:

$$\forall \ell = p+1, \ldots, m, \quad N_{i_\ell} x_{i_\ell} \mathbf{b}_{i_\ell} = N_{j_\ell} y_{i_\ell} \mathbf{b}_{y_\ell}. \tag{27}$$

From the latter equalities we get

$$(\epsilon_1, \ldots, \epsilon_m) = (\delta_1, \ldots, \delta_m),$$

and

$$(r_{i_{p+1}} \cdots r_{i_m}) = (s_{j_{q+1}} \cdots s_{j_m}),$$

thus implying $\mathcal{B}_i = \mathcal{B}_j$, a contradiction.

Let us prove that, for every $j = 1, \ldots, s$, $L^- \cap L'_j = \emptyset$. By contradiction, deny the claim. Hence there exist an index j, $1 \le j \le s$, and a word u such that $u \in L^- \cap L'_j$. Since, by definition of L'_j, u possesses a factor $w \in \mathcal{W}$ and since $L^- \subseteq u_1^* \cdots u_k^*$, (cf. Remark 3) one gets a contradiction. Hence $L^- \cap L'_j = \emptyset$. □

Lemma 10. *Let i be an index with $1 \le i \le s$. Let L'_i and L_i be the corresponding languages of the families (23) and (24), respectively. Then L'_i is commutatively equivalent to L_i.*

Proof. Let i be an index with $1 \le i \le s$. We prove that L_i is commutatively equivalent to L'_i. By (24), one has $L_i = \varphi(\mathcal{B}_i)$ where:

$$\begin{aligned}\mathcal{B}_i = B_{r_{i_{p+1}} \cdots r_{i_m}}^{(\epsilon_1, \ldots, \epsilon_m)} &= \mathbf{b}_0 + \left\{ \sum_{\ell=1}^{p} c_{i_\ell} \mathbf{b}_{i_\ell} \;\middle|\; 0 \le c_{i_\ell} < N_{i_\ell} \right\} \\ &+ \sum_{\ell=p+1}^{m} r_{i_\ell} \mathbf{b}_{i_\ell} + \left\{ \sum_{\ell=p+1}^{m} (N_{i_\ell} x_{i_\ell}) \mathbf{b}_{i_\ell} \;\middle|\; x_{i_\ell} \ge 1 \right\}.\end{aligned}$$

Moreover, by (25), L'_i is the language defined as:

$$L'_i = \varphi\left(\mathbf{b}_0 + \left\{ \sum_{\ell=1}^{p} c_{i_\ell} \mathbf{b}_{i_\ell} \;\middle|\; 0 \le c_{i_\ell} < N_{i_\ell} \right\} + \sum_{\ell=p+1}^{m} r_{i_\ell} \mathbf{b}_{i_\ell} \right) w_{i_{p+1}}^+ w_{i_{p+2}}^+ \cdots w_{i_m}^+.$$

Let $u \in L_i$. By Proposition 2 there exists exactly one vector $\mathbf{b} \in \mathcal{B}_i = B_{r_{i_{p+1}} \cdots r_{i_m}}^{(\epsilon_1, \ldots, \epsilon_m)}$ such that $u = \varphi(\mathbf{b})$. Moreover by the fact that $\mathbf{b}_0, \ldots, \mathbf{b}_m$ form the representation of B, there exist exactly a tuple $x_{i_\ell} \ge 1$, with $i_\ell = p+1, \ldots, m$ and a tuple c_{i_ℓ}, with $i_\ell = 1, \ldots, p$ where $0 \le c_{i_\ell} < N_{i_\ell}$ such that

$$\mathbf{b} = \mathbf{v} + \sum_{\ell=p+1}^{m} N_{i_\ell} x_{i_\ell} \mathbf{b}_{i_\ell}, \quad x_{i_\ell} > 0, \tag{28}$$

with

$$\mathbf{v} = \mathbf{b}_0 + \sum_{\ell=1}^{p} c_{i_\ell} \mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} r_{i_\ell} \mathbf{b}_{i_\ell}. \tag{29}$$

Let us consider the map:

$$f : L_i \longrightarrow L'_i$$

such that, for every $u \in L_i$,

$$f(u) = \varphi(\mathbf{v}) w_{i_{p+1}}^{x_{i_{p+1}}} \cdots w_{i_m}^{x_{i_m}}.$$

The remark above implies that f is well defined as a map from L_i to L'_i. Our main task is to prove that f is a bijection from L_i to L'_i that preserves the Parikh vectors of words of L_i.

Let us prove that f is a bijection from L_i to L'_i. 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 $u' \in L_i$ and assume $f(u) = f(u')$. As before, one has $u' = \varphi(\mathbf{b}')$ where:

$$\mathbf{b}' = \mathbf{v}' + \sum_{\ell=p+1}^{m} N_{i_\ell} y_{i_\ell} \mathbf{b}_{i_\ell}, \quad y_{i_\ell} > 0,$$

with

$$\mathbf{v}' = \mathbf{b}_0 + \sum_{\ell=1}^{p} c'_{i_\ell} \mathbf{b}_{i_\ell} + \sum_{\ell=p+1}^{m} r_{i_\ell} \mathbf{b}_{i_\ell}, \quad 0 \le c'_{i_\ell} < N_{i_\ell}.$$

Then

$$f(u') = \varphi(\mathbf{v}') w_{i_{p+1}}^{y_{i_{p+1}}} \cdots w_{i_m}^{y_{i_m}}.$$

Let $U_1 = w_{i_{p+1}}^{x_{i_{p+1}}} \cdots w_{i_m}^{x_{i_m}}$ and $U_2 = w_{i_{p+1}}^{y_{i_{p+1}}} \cdots w_{i_m}^{y_{i_m}}$. Let us first show that $|U_1| = |U_2|$. By contradiction, deny the claim. Assume $|U_1| < |U_2|$ (similarly one treats the other case). Since $U_1, U_2 \in \mathcal{W}^+$ and all the words of $\mathcal{W}$ have the same length, one has that there exists a word $w \in \mathcal{W}$ which is a suffix of $\varphi(\mathbf{v})$. Since $\varphi(\mathbf{v}) \in u_1^* \cdots u_k^*$, one gets a contradiction (cf. Remark 3). Thus $|U_1| = |U_2|$. Hence we have $\varphi(\mathbf{v}) = \varphi(\mathbf{v}')$ and $U_1 = U_2$. From $U_1 = U_2$, since $\mathcal{W}$ is a code, one has that

$$x_{i_{p+1}} = y_{i_{p+1}}, \qquad x_{i_{p+2}} = y_{i_{p+2}}, \ldots, x_{i_m} = y_{i_m}.$$

On the other hand, from $\varphi(\mathbf{v}) = \varphi(\mathbf{v}')$, by the injectivity of φ over B, one has $\mathbf{v} = \mathbf{v}'$. Hence $\mathbf{b} = \mathbf{b}'$ and thus $u = u'$. Thus f is injective on L_i.

Finally, by (19) and by using Lemma 1, one has that: for every $u \in L_i$,

$$\begin{aligned}
\psi(u) &= \psi\big(\varphi(\mathbf{b})\big) = \psi\left(\varphi\left(\mathbf{v} + \sum_{\ell=p+1}^{m} N_{i_\ell} x_{i_\ell} \mathbf{b}_{i_\ell}\right)\right) \\
&= \psi\big(\varphi(\mathbf{v})\varphi(N_{i_{p+1}} x_{i_{p+1}} \mathbf{b}_{i_{p+1}}) \cdots \varphi(N_{i_m} x_{i_m} \mathbf{b}_{i_m})\big) \\
&= \psi\big(\varphi(\mathbf{v}) w_{i_{p+1}}^{x_{i_{p+1}}} \cdots w_{i_m}^{x_{i_m}}\big) = \psi\big(f(u)\big),
\end{aligned}$$

and this concludes the proof. □

Let L' be the language over A:

$$L' = L^- \cup \bigcup_{j=1}^{s} L'_j, \tag{30}$$

where the languages L'_j, $j = 1, \ldots, s$ are defined in (23).

The following corollary provides an affirmative answer to the CE Problem under Assumption 1.

Corollary 1. *Let L be a bounded context-free language satisfying Assumption 1. Then L' is a regular language commutatively equivalent to L. Moreover, L' can be effectively constructed starting from an effective presentation of L.*

Proof. Since L^- is finite and, for every $j = 1, \dots, s$, $L'_j \in Rat(A^*)$, then one has $L' \in Rat(A^*)$.

By Proposition 4, the languages L_j, $j = 1, \dots, s$, of (24), together with L^-, give a partition of L. On the other hand, by Lemma 9, the languages L'_j, $j = 1, \dots, s$, of (23), together with L^-, give a partition of L'. By Lemma 10, for every $j = 1, \dots, s$, $L_j \sim L'_j$. The claim follows by applying Lemma 2. □

Remark 4. Observe that if the language L is finite, that is, L is described via φ by a finite set of vectors, the claim is trivially achieved.

4.2. The semi-simple case

In this section we prove Theorem 2 in the full generality. Thus we can suppose that L is described, via the Ginsburg map, by a semi-simple set

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

where, $n \geq 2$, B_0 is finite and, for every $i = 1, \dots, n$ B_i is a simple set of dimension k_i with $0 < k_i \leq k$. Moreover, these sets satisfy Assumption 1, that is:

- for every $i = 1, \dots, n$, if B_i is written as:

$$B_i = \left\{ \mathbf{b}_0^{(i)} + x_1 \mathbf{b}_1^{(i)} + \dots + x_{k_i} \mathbf{b}_{k_i}^{(i)} \;\middle|\; x_\ell \in \mathbb{N} \right\},$$

  where the vectors $\mathbf{b}_0^{(i)}, \mathbf{b}_1^{(i)}, \dots, \mathbf{b}_{k_i}^{(i)}$ form the representation of B_i, then, for every $\ell = 1, \dots, k_i$, with $k_i > 0$, the word $\varphi(\mathbf{b}_\ell^{(i)})$ has, at least, two distinct letters in its factorization.

Since the sets B_i, with $i = 0, \dots, n$ are pairwise disjoint and since, by Proposition 2, we may assume that φ is injective on B, for all i, j with $0 \leq i < j \leq n$, we have $\varphi(B_i) \cap \varphi(B_j) = \emptyset$. Then setting, for every $i = 0, \dots, n$, $L_i = \varphi(B_i)$, we have

$$L = \bigcup_{i=0}^{n} L_i,$$

and the languages L_i form a partition of L.

We show the proof in the case $n = 2$ since the general case, *i.e.* for $n \geq 3$, follows the very same scheme.

We apply the algorithm described in Section 4.1 to the simple set B_1. Starting from the vectors $\mathbf{b}_0^{(1)}, \mathbf{b}_1^{(1)}, \dots, \mathbf{b}_{k_1}^{(1)}$, we effectively compute the sequence (21) that we denote:

$$N_1^{(1)}, \dots, N_{k_1}^{(1)}, \beta^{(1)}.$$

Starting from that sequence, we effectively construct the regular language L'_1 (cf. (30)) which is commutatively equivalent to $L_1 = \varphi(B_1)$. We recall that

$$L'_1 = L_1^- \cup \bigcup_{j=1}^{s_1} L'_{1j},$$

where $L_1^- = \varphi(B_1^-)$ and, for every $j = 1, \dots, s_1$, the language L'_{1j} is defined as (cf. (23))

$$L'_{1j} = \varphi(\mathbf{V}_{1j}) w_{i_{p+1}}^+ w_{i_{p+2}}^+ \cdots w_{i_{k_1}}^+,$$

where, for every $\ell = p+1, \dots, k_1$, w_{i_ℓ} are the words of the code (18) and

$$\mathbf{V}_{1j} = \mathbf{b}_0^{(1)} + \left\{ \sum_{\ell=1}^{p} c_{i_\ell} \mathbf{b}_{i_\ell}^{(1)} \;\middle|\; 0 \leq c_{i_\ell} < N_{i_\ell}^{(1)} \right\} + \sum_{\ell=p+1}^{k_1} r_{i_\ell} \mathbf{b}_{i_\ell}^{(1)},$$

where, for every $\ell = p+1, \dots, k_1$, $0 \leq r_{i_\ell} < N_{i_\ell}^{(1)}$.

Now we apply the algorithm described in Section 4.1 to B_2. Starting from the vectors $\mathbf{b}_0^{(2)}, \mathbf{b}_1^{(2)}, \dots, \mathbf{b}_{k_2}^{(2)}$, we effectively compute a sequence (21):

$$N_1^{(2)}, \dots, N_{k_2}^{(2)}, \beta^{(2)},$$

where $\beta^{(2)}$ is a positive multiple of $\beta^{(1)}$.

Starting from the sequence $(N_1^{(2)}, \ldots, N_{k_2}^{(2)})$, we effectively construct a regular language L'_2 (cf. (30)) which is commutatively equivalent to $L_2 = \varphi(B_2)$ and such that

$$L'_2 = L_2^- \cup \bigcup_{j=1}^{s_2} L'_{2j},$$

where $L_2^- = \varphi(B_2^-)$ and, for every $j = 1, \ldots, s_2$, the language L'_{2j} is defined as (cf. (23))

$$L'_{2j} = \varphi(\mathbf{V}_{2j}) v_{i_{q+1}}^+ v_{i_{q+2}}^+ \cdots v_{i_{k_2}}^+,$$

where, for every $\ell = q+1, \ldots, k_2$, v_{i_ℓ} are the words of the code (18) and

$$\mathbf{V}_{2j} = \mathbf{b}_0^{(2)} + \left\{ \sum_{\ell=1}^{q} d_{i_\ell} \mathbf{b}_{i_\ell}^{(2)} \,\middle|\, 0 \le d_{i_\ell} < N_{i_\ell}^{(2)} \right\} + \sum_{\ell=q+1}^{k_2} t_{i_\ell} \mathbf{b}_{i_\ell}^{(2)},$$

where, for every $\ell = q+1, \ldots, k_2$, $0 \le t_{i_\ell} < N_{i_\ell}^{(2)}$.

The following lemmata hold.

Lemma 11. *$L_1^- \cap L_2^- = \emptyset$ and $L_0 \cap L_i^- = \emptyset$ for $i = 1, 2$.*

Proof. Let us prove the first equality. By contradiction, deny the claim so that there exists a word $u \in L_1^- \cap L_2^-$. Thus, $u = \varphi(\mathbf{b}) = \varphi(\mathbf{b}')$ where $\mathbf{b} \in B_1, \mathbf{b}' \in B_2$. Since φ is injective on B and $B_1 \cap B_2 = \emptyset$, one gets a contradiction. Hence $L_1^- \cap L_2^- = \emptyset$. The second part of the claim is proved by using the same argument. □

Lemma 12. *Let $i, j \in \mathbb{N}$ with $1 \le i \le s_1$ and $1 \le j \le s_2$. Then one has*

(i) $L'_{1j} \cap L_2^- = L'_{1j} \cap L_0 = \emptyset$;
(ii) $L'_{2j} \cap L_1^- = L'_{2j} \cap L_0 = \emptyset$;
(iii) $L'_{2j} \cap L'_{1i} = \emptyset$.

Proof. (i) Let us prove $L'_{1j} \cap L_2^- = \emptyset$. The argument is the same used in the second part of Lemma 9. By contradiction, deny the claim. Hence there exist an index j, $1 \le j \le s_1$, and a word u such that $u \in L_2^- \cap L'_{1j}$. Since, by definition of L'_{1j}, u possesses a factor $w \in \mathcal{W}$ and since $L_2^- \subseteq u_1^* \cdots u_k^*$, one gets a contradiction (cf. Remark 3). The second equality is proved by using the same argument. (ii) is proved similarly as (i).

Let us finally prove (iii). By contradiction, deny the claim. Hence there exist indices i, j, with $1 \le i \le s_1$ and $1 \le j \le s_2$ such that $L'_{2j} \cap L'_{1i} \ne \emptyset$. Let $u \in L'_{2j} \cap L'_{1i}$. Since $u \in L'_{1i}$, one has:

$$u = \varphi(\mathbf{v}) w_{i_{p+1}}^{x_{i_{p+1}}} \cdots w_{i_{k_1}}^{x_{i_{k_1}}}, \quad \mathbf{v} \in \mathbf{V}_{1i},$$

and, since $u \in L'_{2j}$, one has:

$$u = \varphi(\mathbf{v}') v_{i_{q+1}}^{y_{i_{q+1}}} \cdots v_{i_{k_2}}^{y_{i_{k_2}}}, \quad \mathbf{v}' \in \mathbf{V}_{2j},$$

whence

$$\varphi(\mathbf{v}) w_{i_{p+1}}^{x_{i_{p+1}}} \cdots w_{i_{k_1}}^{x_{i_{k_1}}} = \varphi(\mathbf{v}') v_{i_{q+1}}^{y_{i_{q+1}}} \cdots v_{i_{k_2}}^{y_{i_{k_2}}}. \tag{31}$$

One easily checks that $|\varphi(\mathbf{v})| = |\varphi(\mathbf{v}')|$. Indeed, assume $|\varphi(\mathbf{v}')| < |\varphi(\mathbf{v})|$, the other case being completely similar. This implies that the number $|\varphi(\mathbf{v})| - |\varphi(\mathbf{v}')|$ is a positive multiple of $\beta^{(1)}$. By (31), the prefix of length $\beta^{(1)}$ of $v_{i_{q+1}}$ is a factor of $\varphi(\mathbf{v})$. Since $\varphi(\mathbf{v}) \in u_1^* \cdots u_k^*$, this implies that the prefix of length $\beta^{(1)}$ of $v_{i_{q+1}}$ is a factor of a word of $u_1^* \cdots u_k^*$ which is not possible (cf. Remark 3). Hence $|\varphi(\mathbf{v})| = |\varphi(\mathbf{v}')|$ so that $\varphi(\mathbf{v}) = \varphi(\mathbf{v}')$. Since $\mathbf{v}, \mathbf{v}' \in B$ and φ is injective on B, one has $\mathbf{v} = \mathbf{v}'$. By observing that $\mathbf{v} \in B_1$ and $\mathbf{v}' \in B_2$, one has $\mathbf{v} \in B_1 \cap B_2$, so $B_1 \cap B_2 \ne \emptyset$, which is a contradiction. Condition (iii) is therefore proved. □

As an immediate consequence of Lemma 11 and Lemma 12, we have

Corollary 2. *The languages L_0, L'_1 and L'_2 are pairwise disjoint.*

We are now able to prove Theorem 2.

Proof. If L is finite, the claim is trivially proved. Hence suppose that L is not finite. By applying the previous construction to L, we get a regular language $L' \in Rat(A^*)$:

$$L' = L_0 \cup L'_1 \cup L'_2.$$

By Corollary 2 and by taking into account that

$$L = L_0 \cup L_1 \cup L_2,$$

and

$$L'_1 \sim L_1, \qquad L'_2 \sim L_2$$

the fact that $L \sim L'$ follows by applying Lemma 2. Since all the steps of the proof are constructive, L' can be effectively constructed starting from an effective presentation of L. $\square$

Acknowledgements

We thank the anonymous referee for her (his) help in improving a previous version of this paper.

Appendix A

Proof of Lemma 3. Let $u_1, \ldots, u_k$ be the sequence of words of the lemma. Let a be any letter of A. Consider any word of the form $u_j u_{j+1}$ such that both u_j and u_{j+1} have at least one occurrence of a letter different from a in their factorization. Then there exists a constant γ_j such that for every word u' of u_j^*, or of u_{j+1}^*, or of $(u_j u_{j+1})^*$, every factor w' of u', with $w' \in a^*$ has length less than γ_j. Indeed, it is enough to set $\gamma_j = |u_j| + |u_{j+1}|$. Now let $\gamma_a = 1 + \max \gamma_j$, as j varies on all the above considered words and define γ as the greatest of all constants γ_a as a varies on A. We claim that γ has the required property. Arguing by contradiction, let $a, b \in A$, with $a \neq b$, be fixed and consider a word w having a factor of the form

$$\underbrace{a^\gamma b a^\gamma b \ldots a^\gamma b a^\gamma}_{(k+1)\text{-times}} \tag{32}$$

where a^γ occurs $(k+1)$-times in the word (32).

Assume w is a factor of a word in $u_1^* \cdots u_k^*$. Now by the choice of γ, every a^γ cannot be a factor of any power of a word u_j or u_{j+1} or $u_j u_{j+1}$, with $1 \leq j < k$, if both u_j and u_{j+1} have at least one occurrence of a letter different from a in their factorization. Therefore a^γ must be a factor of a word u_i^t, for some $t \in \mathbb{N}$, with u_i occurring in the sequence $u_1, \ldots, u_k$ and such that $u_i \in a^*$. Moreover, as such factors a^γ are separated by an occurrence of b, every factor must belong to a power of a different word u_i. But now observe that such words u_i are no more than k, whilst there are at least $k+1$ such factors in w and we have reached a contradiction. $\square$

Example 1. Let $A = \{a, b\}$ and $\psi : A^* \longrightarrow \mathbb{N}^2$. Thus for every $u \in A^*$, one has $\psi(u) = (|u|_a, |u|_b)$. Let $u_1^* u_2^* = (ab)^*(ba)^*$ where $u_1 = ab, u_2 = ba$. The forbidden word (3) of Lemma 3 is $a^2ba^2ba^2$. Let us consider the list of words

$$a^{36}b^{72}, \quad a^{36}b^{72}, \quad b^{30}ab^2ab^4a^{70}, \quad (ba)^{10}(ab)^{11}b^{15}a^{51},$$

where the corresponding multiset of Parikh vectors is

$$\{(2, 36(1,2)), (2, 36(2,1))\}.$$

Then one has:

$$\mathcal{W} = \mathcal{W}_{(1,2)} \cup \mathcal{W}_{(2,1)},$$

where

$$\mathcal{W}_{(1,2)} = \{a^2ba^2ba^2a^{66}b^{34}, a^2ba^2ba^2ba^{66}b^{33}\},$$

and

$$\mathcal{W}_{(2,1)} = \{a^2ba^2ba^2a^{30}b^{70}, a^2ba^2ba^2ba^{30}b^{69}\}.$$

Example 2. Let $B = \{x_1\mathbf{b}_1 + x_2\mathbf{b}_2 : x_1, x_2 \in \mathbb{N}\}$ be a simple set of dimension 2 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 $N_1 = 2$, $N_2 = 3$. Then the partition of B is:

$$B = B^- \cup B^{(-,+)} \cup B^{(+,-)} \cup B^{(+,+)},$$

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 instance, if $(\epsilon_1, \epsilon_2) = (+,+)$ and $r_1 = 1$, $r_2 = 2$, one has:

$$B_{1,2}^{(+,+)} = \big\{(\mathbf{b}_1 + 2\mathbf{b}_2) + 2x_1\mathbf{b}_1 + 3x_2\mathbf{b}_2 : x_1, x_2 \geq 1\big\},$$

and, if $(\epsilon_1, \epsilon_2) = (-,+)$ and $r_2 = 1$, one has:

$$\begin{aligned} B_1^{(-,+)} &= \bigcup_{0 \leq c_1 < 2} \{c_1\mathbf{b}_1 + \mathbf{b}_2 + 3x_2\mathbf{b}_2 : x_2 \geq 1\} \\ &= \{\mathbf{b}_2 + 3x_2\mathbf{b}_2 : x_2 \geq 1\} \cup \{\mathbf{b}_1 + \mathbf{b}_2 + 3x_2\mathbf{b}_2 : x_2 \geq 1\}. \end{aligned}$$

References

[1] M.-P. Béal, D. Perrin, On the generating sequences of regular languages on k symbols, J. ACM 50 (2003) 955–980.

[2] J. Berstel, D. Perrin, C. Reutenauer, Codes and Automata, Encyclopedia of Mathematics and Its Applications, vol. 129, Cambridge University Press, Cambridge, 2009.

[3] L. Boasson, A. Restivo, Une caractérisation des langages algébriques Bornés, RAIRO Inform. Théor. 11 (1977) 203–205.

[4] N. Chomsky, M.-P. Schützenberger, The algebraic theory of context-free languages, in: P. Braffort, D. Hirschberg (Eds.), Computer Programming and Formal Systems, North Holland Publishing Company, Amsterdam, 1963, pp. 118–161.

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

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

[7] 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, in: Electronic Proceedings in Theoretical Computer Science, 2011, pp. 1–21, http://dx.doi.org/10.4204/EPTCS.63.

[8] F. D'Alessandro, B. Intrigila, S. Varricchio, Quasi-polynomials, semi-linear set, and linear diophantine equations, Theoret. Comput. Sci. 416 (2012) 1–16.

[9] F. D'Alessandro, B. Intrigila, On the commutative equivalence of semi-linear sets of $\mathbb{N}^k$, Theoret. Comput. Sci. 562 (2015) 476–495, this issue.

[10] F. D'Alessandro, B. Intrigila, On the commutative equivalence of bounded context-free and regular languages: the semi-linear case, preprint (2013), submitted to Theoret. Comput. Sci.

[11] A. de Luca, S. Varricchio, Finiteness and Regularity in Semigroups and Formal Languages, Springer-Verlag, Berlin, 1999.

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

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

[14] S. Ginsburg, The Mathematical Theory of Context-Free Languages, McGraw-Hill, New York, 1966.

[15] S. Ginsburg, E.H. Spanier semigroups, Presburger formulas, and languages, Pacific J. Math. 16 (1966) 285–296.

[16] J. Honkala, A decision method for Parikh slenderness of context-free languages, Discrete Appl. Math. 73 (1997) 1–4.

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

[18] J. Honkala, On Parikh slender context-free languages, Theoret. Comput. Sci. 255 (2001) 667–677.

[19] O. Ibarra, B. Ravikumar, On sparseness, ambiguity and other decision problems for acceptors and transducers, in: Lecture Notes in Computer Science, vol. 210, Springer-Verlag, Berlin, 1986, pp. 171–179.

[20] L. Ilie, G. Rozenberg, A. Salomaa, A characterization of poly-slender context-free languages, RAIRO Inform. Théor. Appl. 34 (2000) 77–86.

[21] R. Ito, Every semi-linear set is a finite union of disjoint linear sets, J. Comput. System Sci. 3 (1969) 295–342.

[22] J. Kortelainen, T. Salmi, There does not exist a minimal full trio with respect to bounded context-free languages, in: Lecture Notes in Computer Science, vol. 6795, Springer-Verlag, Berlin, 2011, pp. 312–323.

[23] M. Latteux, G. Thierrin, On bounded context-free languages, Elektron. Inform. Verarb. Kybern. 20 (1984) 3–8.

[24] R.J. Parikh, On context-free languages, J. ACM 13 (1966) 570–581.

[25] A. Restivo, A characterization of bounded regular sets, in: Lecture Notes in Computer Science, vol. 33, Springer-Verlag, Berlin, 1975, pp. 239–244.