Measuring Power of Commutative Group Languages

Takao Yuyama[1] and Ryoma Sin'ya[2]

[1] Research Institute for Mathematical Sciences, Kyoto University, Kyoto, Japan
yuyama@kurims.kyoto-u.ac.jp
[2] Akita University, Akita, Japan
ryoma@math.akita-u.ac.jp

Abstract. A language L is said to be $\mathcal{C}$-measurable, where $\mathcal{C}$ is a class of languages, if there is an infinite sequence of languages in $\mathcal{C}$ that "converges" to L. In this paper, we investigate the measuring powers of Gcom of the class of all languages recognised by finite commutative groups and its subclass named MOD. A language is in MOD if membership of a word in the language only depends on its length modulo some fixed integer. In particular, we show that, for a given regular language L, it is decidable whether L is Gcom-measurable (MOD-measurable, respectively) or not. Our results demonstrate that there is a huge gap between the expressive power of group languages and commutative group languages, even from a (very rough) measure theoretic point of view.

1 Introduction

The notion of $\mathcal{C}$-measurability for a class $\mathcal{C}$ of languages is introduced by [6] and it was used for classifying non-regular languages by using regular languages. A language L is said to be $\mathcal{C}$-measurable if there is an infinite sequence of languages in $\mathcal{C}$ that converges to L. Also, the $\mathcal{C}$-measurability can be defined by using so-called *Carathéodory extension* [7], a purely measure theoretic notion. Roughly speaking, L is $\mathcal{C}$-measurable means that it can be approximated by a language in $\mathcal{C}$ with arbitrary high precision: the notion of "precision" is formally defined by the *density* of formal languages. Hence that a language L is not $\mathcal{C}$-measurable ($\mathcal{C}$-*im*measurable) means that L has a complex shape so that it can not be approximated by languages in $\mathcal{C}$.

So far, the decidability and different characterisation of $\mathcal{C}$-measurability is systematically studied for some subclass $\mathcal{C}$ of star-free languages. For example, two simple and decidable characterisations of PT-measurable and AT-measurable languages are given in [8], where PT is the class of all piecewise testable languages and AT is the class of all alphabet testable languages (*i.e.*, languages definable by the first-order logic with only one variable). While the AT-measurability is strictly weaker than PT-measurability, it is interesting that the computational complexity of the AT-measurability is much higher than PT-measurability: deciding the AT-measurability of a given regular language L (represented by a deterministic automaton) is *PSPACE-complete* but deciding the PT-measurability of L can be done *in linear time* with respect to the number of states [10]. Furthermore, in [8,9] it was shown that the GD-measurability and UPol-measurability are *equivalent*, where GD is the class of all generalised definite languages (*i.e.*, languages that can be defined by a finite Boolean combination of prefix and suffix tests of some bounded length) and UPol is the class of all unambiguous polynomials (*i.e.*, languages definable by the first-order logic with only two variables). Some properties of the measuring power of star-free languages is investigated in [7], but the decidability for regular languages is still unknown.

On the other hand, for the measuring power of *group languages*, nothing is yet known. A language is said to be a group language if its syntactic monoid is a finite group, equivalently, it can be recognised by a permutation automaton [11]. Although the definition of group languages is very simple from an algebraic point of view, this class dose not have a language theoretic nor logical different characterisation. To borrow the phrase used by Place–Zeitoun [4]: "This makes it difficult to get an intuitive grasp about group languages, which may explain why this class remains poorly understood."

This paper addresses the first development step for understanding the measuring power of group languages. In this paper, we investigate the measuring power of G the class of all group languages, Gcom the class of commutative group languages and MOD a subclass of Gcom (see Section 2 for the precise definition). The main results of this paper are three kinds.

(1) Every group language whose syntactic monoid is non-commutative is Gcom-*im*measurable (Theorem 1).
(2) A decidable characterisation of the Gcom-measurability for regular languages (Theorem 2).
(3) A simple decidable characterisation of the MOD-measurability (Theorem 3).

2 Preliminaries

This section provides the precise definitions of density, measurability, group languages and its two subclasses. REG_A denotes the family of all regular languages over an alphabet A. For a word $w \in A^*$ and a letter $a \in A$, we write $|w|_a$ the number of occurrences of a in w. The complement of L over A is denoted by $L^{\mathrm{c}} = A^* \setminus L$. We assume that the reader has a standard knowledge of algebraic language theory (*cf.* [2,3]).

The main targets of this paper are the following three subclasses of regular languages. The *group languages* G, the *commutative group languages* Gcom and the *modulo languages* MOD:

$$\mathsf{G}_A \stackrel{\mathsf{def}}{=} \{\, L \subseteq A^* \mid L \text{ is recognised by a finite group} \,\}$$

$$\mathsf{Gcom}_A \stackrel{\mathsf{def}}{=} \{\, L \subseteq A^* \mid L \text{ is recognised by a finite commutative group} \,\}$$

$$\mathsf{MOD}_A \stackrel{\mathsf{def}}{=} \left\{ L \subseteq A^* \;\middle|\; \begin{array}{l} L \text{ is recognised by a morphism } \eta\colon A^* \to G \text{ into a} \\ \text{finite group } G \text{ such that } \eta(a) = \eta(b) \text{ for all } a, b \in A \end{array} \right\}$$

A monoid M divides a monoid N if M is a monoid homomorphic image of a submonoid of N. It is well-known that a monoid M recognises a language L if and only if the syntactic monoid M_L of L divides M (*cf.* [2, Theorem 10.2.6]), hence L is in G_A (Gcom_A, respectively) if and only if M_L is a finite group (a finite commutative group, respectively). For $a \in A$, $q, r \in \mathbb{N}$ such that $r < q$, we let

$$L^a_{q,r} = \{\, w \in A^* \mid |w|_a \equiv r \mod q \,\} \text{ and } L_{q,r} = \{\, w \in A^* \mid |w| \equiv r \mod q \,\}.$$

It is also known that Gcom_A (MOD_A, respectively) is the smallest Boolean algebra containing all languages of $L^a_{q,r}$ ($L_{q,r}$, respectively) (*cf.* [3,4]). Hereafter, we only consider an alphabet A such that $\#(A) \geq 2$ (because $\mathsf{MOD}_A = \mathsf{Gcom}_A = \mathsf{G}_A$ if $\#(A) = 1$) and sometimes omit the subscript A for denoting these classes of languages.

2.1 Density and measurability of formal language

For a set X, we denote by $\#(X)$ the cardinality of X. We denote by $\mathbb{N}$ the set of natural numbers including 0.

Definition 1 (*cf.* [1]). The *density* $\delta_A(L)$ of $L \subseteq A^*$ is defined as

$$\delta_A(L) \overset{\mathtt{def}}{=\!=} \lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} \frac{\#(L \cap A^k)}{\#(A^k)}$$

if it exists, otherwise we write $\delta_A(L) = \bot$. The language L is called *null* if $\delta_A(L) = 0$, and dually, L is called *co-null* if $\delta_A(L) = 1$.

Example 1. It is known that every regular language has a rational density (*cf.* [1]) and it is computable. For each word w, the language A^*wA^*, the set of all words that contain w as a factor, is of density one (co-null). This fact follows from the so-called *infinite monkey theorem*: take any word w. A random word of length n contains w as a factor with probability tending to 1 as $n \to \infty$.

We list some basic properties of the density as follows.

Lemma 1. *Let $K, L \subseteq A^*$ with $\delta_A(K) = \alpha, \delta_A(L) = \beta$. Then we have: (1) $\delta_A(L \setminus K) = \beta - \alpha$ if $K \subseteq L$. (2) $\delta_A(K^{\mathrm{c}}) = 1 - \alpha$. (3) $\delta_A(K \cup L) = \alpha + \beta$ if $K \cap L = \varnothing$.*

Lemma 2. *Let $\eta \colon A^* \to G$ be a morphism onto a finite group. For any $g \in G$, we have $\delta_A(\eta^{-1}(g)) = \#(G)^{-1}$.*

For more detailed properties of δ_A, see Chapter 13 of [1].

The notion of "measurability" on formal languages is defined by a standard measure theoretic approach as follows.

Definition 2 ([6]). Let $\mathcal{C}_A$ be a family of languages over A. For a language $L \subseteq A^*$, we define its $\mathcal{C}_A$*-inner-density* $\underline{\mu}_{\mathcal{C}_A}(L)$ and $\mathcal{C}_A$*-outer-density* $\overline{\mu}_{\mathcal{C}_A}(L)$ over A as

$$\underline{\mu}_{\mathcal{C}_A}(L) \stackrel{\mathsf{def}}{=} \sup\{\, \delta_A(K) \mid K \subseteq L, K \in \mathcal{C}_A, \delta_A(K) \neq \bot \,\} \text{ and}$$
$$\overline{\mu}_{\mathcal{C}_A}(L) \stackrel{\mathsf{def}}{=} \inf\{\, \delta_A(K) \mid L \subseteq K, K \in \mathcal{C}_A, \delta_A(K) \neq \bot \,\}, \text{ respectively.}$$

A language L is said to be $\mathcal{C}_A$*-measurable* if $\underline{\mu}_{\mathcal{C}_A}(L) = \overline{\mu}_{\mathcal{C}_A}(L)$ holds. We say that an infinite sequence $(L_n)_n$ of languages over A *converges to* L *from inner (from outer,* respectively*)* if $L_n \subseteq L$ ($L_n \supseteq L$, respectively) for each n and $\lim_{n\to\infty} \delta_A(L_n) = \delta_A(L)$.

The following is an example of Gcom-measurable non-regular language.

Example 2 ([6]). The semi-Dyck language $D = \{\varepsilon, ab, aabb, abab, aaabbb, \ldots\}$ over $A = \{a, b\}$ is Gcom-measurable. For each $k \geq 2$, the language $L_k = \{w \in A^* \mid |w|_a = |w|_b \mod k\}$ is in Gcom, because $\eta^{-1}(0) = L_k$ holds for the morphism $\eta\colon A^* \to \mathbb{Z}/k\mathbb{Z}$ where $h(a) = 1$ and $h(b) = k-1$. Obviously, $D \subseteq L_k$ holds and it follows from Lemma 2 that $\delta_A(L_k) = 1/k$ holds. Hence $\delta_A(L_k)$ tends to zero if k tends to infinity.

For a family $\mathcal{C}_A$ of languages over A, we denote by $\mathrm{Ext}_A(\mathcal{C}_A)$ ($\mathrm{RExt}_A(\mathcal{C}_A)$, respectively) the class of all $\mathcal{C}_A$-measurable languages ($\mathcal{C}_A$-measurable *regular* languages, respectively) over A.

Lemma 3 ([7]). *The operator* Ext_A *is a closure, i.e., it satisfies the following three properties for each* $\mathcal{C} \subseteq \mathcal{D} \subseteq 2^{A^*}$*: (extensive)* $\mathcal{C} \subseteq \mathrm{Ext}_A(\mathcal{C})$*, (monotone)* $\mathrm{Ext}_A(\mathcal{C}) \subseteq \mathrm{Ext}_A(\mathcal{D})$*, and (idempotent)* $\mathrm{Ext}_A(\mathrm{Ext}_A(\mathcal{C})) = \mathrm{Ext}_A(\mathcal{C})$*. Moreover,* $\mathrm{Ext}_A(\mathcal{C}_A)$ *is closed under Boolean operations and quotients if* $\mathcal{C}_A$ *is closed under Boolean operations and quotients.*

2.2 Semilinear sets and some background from group theory

Let $A = \{a_1, \ldots, a_d\}$. The *Parikh mapping* $\mathrm{Pkh}_A\colon A^* \to \mathbb{N}^d$ is defined by $\mathrm{Pkh}_A(w) \stackrel{\mathsf{def}}{=} (|w|_{a_1}, \ldots, |w|_{a_d})$. This can be naturally extended to a map taking a language L over A: $\mathrm{Pkh}_A(L) = \{\, \mathrm{Pkh}_A(w) \mid w \in L \,\}$. A set $S \subseteq \mathbb{N}^d$ is called *linear* if S is of the form

$$S = \{\, \boldsymbol{c} + x_1\boldsymbol{p}_1 + \cdots + x_k\boldsymbol{p}_k \mid x_i \in \mathbb{N} \text{ for each } i \,\}$$

for some $k \in \mathbb{N}$ and some vectors $\boldsymbol{c}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_k \in \mathbb{N}^d$. In this case, we call $\boldsymbol{c}$ a constant vector and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_k$ are period vectors of (this representation of) S. A set $S \subseteq \mathbb{N}^d$ is called *semilinear* if it is a finite union of linear sets. It is well-known that (i) every regular language has a semilinear Parikh image, and (ii) the complement of a semilinear set is also semilinear (*cf.* [5]).

Definition 3 (span). *For each $\mathbb{K} \in \{\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}\}$ and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r \in \mathbb{K}^r$, we define*

$$\operatorname{span}_{\mathbb{K}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r) \stackrel{\text{def}}{=} \{\, x_1\boldsymbol{p}_1 + \cdots + x_r\boldsymbol{p}_r \mid x_1, \ldots, x_r \in \mathbb{K} \,\}.$$

With this notation, a linear set $S \subseteq \mathbb{N}^d$ is written as $S = \boldsymbol{c} + \operatorname{span}_{\mathbb{N}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r)$.

A finitely generated commutative group M is called *free* if M is isomorphic to $\mathbb{Z}^r$ for some $r \in \mathbb{N}$. This r is called the *rank* of M, denoted by $\operatorname{rank}(M)$. It is known that every subgroup M of $\mathbb{Z}^r$ is also free of rank $\leq r$.

Proposition 1. *Let $r \in \mathbb{N}$ and $M \subseteq \mathbb{Z}^r$ be a subgroup such that $\operatorname{rank}(M) = r$. Then there exists $N \in \mathbb{N}$ such that $N\mathbb{Z}^r \subseteq M$. In particular, the index $(\mathbb{Z}^r : M)$ is finite.*

Proof. Let $\{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r\}$ be the standard basis of $\mathbb{Z}^r$. Since $\operatorname{rank}(M) = r$, there exists a free basis $\{\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r\}$ of M such that $M = \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r)$. If we put $M_{\mathbb{Q}} = \operatorname{span}_{\mathbb{Q}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_r)$, then $\dim_{\mathbb{Q}}(M_{\mathbb{Q}}) = r$, *i.e.*, $M_{\mathbb{Q}} = \mathbb{Q}^r \supseteq \mathbb{Z}^r$. Therefore there exist $c_{i,j} \in \mathbb{Z}$ $(1 \leq i, j \leq r)$ and $N \in \mathbb{N}$ such that

$$\boldsymbol{e}_i = \frac{c_{i,1}}{N}\boldsymbol{p}_1 + \cdots + \frac{c_{i,r}}{N}\boldsymbol{p}_r$$

for each $i = 1, \ldots, r$. Thus $N\mathbb{Z}^r = \operatorname{span}_{\mathbb{Z}}(N\boldsymbol{e}_1, \ldots, N\boldsymbol{e}_r) \subseteq M$ and $(\mathbb{Z}^r : M) \leq (\mathbb{Z}^r : N\mathbb{Z}^r) = N^r < \infty$. ☐

Definition 4 (rank and index of coset). *For every linear set $\Lambda \subseteq \mathbb{Z}^r$, there exist $\boldsymbol{c} \in \mathbb{Z}^r$ and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\mu \in \mathbb{Z}^r$ such that $\Lambda = \boldsymbol{c} + \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\mu)$. Then $M = \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_\mu)$ is a subgroup of $\mathbb{Z}^r$, and we write*

$$\operatorname{rank}(\Lambda) \stackrel{\text{def}}{=} \operatorname{rank}(M), \qquad (\mathbb{Z}^r : \Lambda) \stackrel{\text{def}}{=} (\mathbb{Z}^r : M).$$

Note that $\operatorname{rank}(\Lambda)$ and $(\mathbb{Z}^r : \Lambda)$ do not depend on the choice of $\boldsymbol{c}, \boldsymbol{p}_1, \ldots, \boldsymbol{p}_\mu$. More generally, we write $(G : gH) \stackrel{\text{def}}{=} (G : H)$ for a group G, a subgroup H, and an element $g \in G$.

Lemma 4. *Let G be a group, $H, K \subseteq G$ be subgroups, and $a, b \in G$. If $aH \cap bK \neq \varnothing$, then $aH \cap bK = x(H \cap K)$ for any $x \in aH \cap bK$.*

Proof. If $x \in aH \cap bK$, then $xH = aH$ and $xK = bK$. Thus $aH \cap bK = xH \cap xK = x(H \cap K)$. ☐

Thanks to Lemma 4, we can write $\operatorname{rank}(\Lambda_1 \cap \Lambda_2)$ consistently.

Proposition 2. *Let $r \in \mathbb{N}$, $M \subseteq \mathbb{Z}^r$ be a subgroup, and $\pi_{n!} \colon \mathbb{Z}^r \to (\mathbb{Z}/n!\mathbb{Z})^r$ be the natural surjection. Then*

$$\operatorname{rank}(M) < r \implies \lim_{n \to \infty} ((\mathbb{Z}/n!\mathbb{Z})^r : \pi_{n!}(M)) = \infty.$$

For the proof of Proposition 2, see Appendix.

For a group G, the *commutator* of two elements $x, y \in G$ is defined as $[x, y] \stackrel{\text{def}}{=} xyx^{-1}y^{-1}$. Note that $[x, y]$ is the identity element if and only if x and y commute. The *commutator subgroup* of G, denoted by $[G, G]$, is the subgroup generated by all the commutators of G. Note that $[G, G]$ is a normal subgroup of G since $z[x, y]z^{-1} = [zxz^{-1}, zyz^{-1}]$ for all $x, y, z \in G$. The *abelianisation* of G is defined as the quotient group $G^{\text{ab}} \stackrel{\text{def}}{=} G/[G, G]$.

3 (Counter)examples of **Gcom**-measurable and **MOD**-measurable languages

To grasp an intuition of Gcom-measurability and MOD-measurability, first we examine concrete examples of Gcom-measurable and MOD-measurable languages.

Proposition 3. *Let* FIN *be the family of all finite and co-finite languages over* A*,* Com *be the all languages whose syntactic monoid is a finite commutative monoid, and* AT *be the Boolean algebra generated by the languages of the form* A^*aA^**.*

(1) $\mathrm{Ext}_A(\mathsf{FIN}) \subsetneq \mathrm{Ext}_A(\mathsf{MOD})$.
(2) $\mathrm{Ext}_A(\mathsf{AT}) \subsetneq \mathrm{Ext}_A(\mathsf{Gcom})$.
(3) $\mathrm{Ext}_A(\mathsf{Gcom}) = \mathrm{Ext}_A(\mathsf{Com})$.

Proof. We first show the strictness of (1) and (2). Every language in $\mathrm{Ext}_A(\mathsf{FIN})$ or $\mathrm{Ext}_A(\mathsf{AT})$ is null or co-null since every language in FIN or AT is so. The language $L_{2,0}$ of even-length words belongs to MOD and has density $\delta_A(L_{2,0}) = 1/2$ by Lemma 2. Thus $L_{2,0} \in \mathrm{Ext}_A(\mathsf{MOD}) \setminus \mathrm{Ext}_A(\mathsf{FIN})$ and $L_{2,0} \in \mathrm{Ext}_A(\mathsf{Gcom}) \setminus \mathrm{Ext}_A(\mathsf{AT})$.

(1) Let $F \subseteq A^*$ be a non-empty finite language. Define $\ell_F = \{\, |w| \mid w \in F \,\} \subseteq \mathbb{N}$ and $N = \max \ell_F$. For each $k \geq 1$, the language

$$F_k = \{\, w \in A^* \mid (|w| \bmod N + k) \in \ell_F \,\}$$

clearly belongs to MOD. By construction, it is easy to see that $F \subseteq F_k$ and $\delta_A(F_k) = \#(\ell_F)\,/(N + k)$ holds. This means that the density of F_k tends to zero if k tends to infinity, *i.e.*, F_k converges to F from outer. Since MOD is closed under Boolean operations, we have $\mathrm{Ext}_A(\mathsf{FIN}) \subseteq \mathrm{Ext}_A(\mathsf{MOD})$ by Lemma 3.

(2) Let $a \in A$ and $B = A \setminus \{a\}$. First we show the Gcom-measurability of $B^* = (A^*aA^*)^{\mathrm{c}} \in \mathsf{AT}$. The construction of approximations is similar with the proof of Item (1). For each $k \geq 2$, the language $L^a_{k,0} = \{\, w \in A^* \mid |w|_a \equiv 0 \mod k \,\}$ is a commutative group language as stated in Section 2. Clearly, $L^a_{k,0}$ satisfies $B^* \subseteq L_k$ and it is easy to see that $\lim_{k\to\infty} \delta_A(L_k) = 0$ holds. Also, Gcom is closed under Boolean operations hence we have $\mathrm{Ext}_A(\mathsf{AT}) \subseteq \mathrm{Ext}_A(\mathsf{Gcom})$ by Lemma 3.

(3) The inclusion $\mathrm{Ext}_A(\mathsf{Gcom}) \subseteq \mathrm{Ext}_A(\mathsf{Com})$ is clear because $\mathsf{Gcom} \subseteq \mathsf{Com}$ holds and by the monotonicity of Ext_A (Lemma 3). To show the reverse inclusion, it is enough to show that $\mathsf{Com} \subseteq \mathrm{Ext}_A(\mathsf{Gcom})$ holds thanks to the idempotency of Ext_A (Lemma 3). We use the following fact [3, Proposition 1.11]: Com is the Boolean algebra generated by the languages of the form $L(a, r) = \{\, w \in A^+ \mid |w|_a = r \,\}$ and $L^a_{q,r}$ where $a \in A$ and $0 \leq r < q$. The Gcom-measurability of $L(a, r)$ can be shown by the same manner with the proof of Item (2). Hence we have $\mathrm{Ext}_A(\mathsf{Gcom}) = \mathrm{Ext}_A(\mathsf{Com})$ by Lemma 3.

□

If we want to show the $\mathcal{C}$-measurability of a given language, the tactics is rather clear: to create a convergent sequence of languages in $\mathcal{C}$. However, to show the $\mathcal{C}$-*im*measurability of a given language, there is no routine tactics and it is much harder in most cases. The following theorem gives infinitely many non-trivial examples of Gcom-immeasurable languages.

Theorem 1. *Every group language whose syntactic monoid is non-commutative is* Gcom-im*measurable.*

Proof. Let $L \in \mathsf{G}_A$ and suppose that the syntactic monoid $G = M_L$ is a non-commutative group. Let $\alpha\colon A^* \to G$ be the canonical surjection. Since G is non-commutative, the commutator subgroup $[G, G]$ of G is nontrivial, *i.e.*, $\{1\} \subsetneq [G, G] \subseteq G$. Hence the abelianisation $G^{\mathrm{ab}} = G/[G, G]$ of G satisfies $\#(G) > \#\big(G^{\mathrm{ab}}\big)$. Let $\pi\colon G \to G^{\mathrm{ab}}$ be the natural surjection. Then L is not recognised by $\pi \circ \alpha$ (otherwise G *divides* G^{ab}, a contradiction). That is, there exist two words $u, v \in A^*$ such that $u \in L \not\ni v$ and $\pi \circ \alpha(u) = \pi \circ \alpha(v)$. Hence we have $\alpha(u)^{-1}\alpha(v) \in [G, G]$, *i.e.*, there exist $x_1, y_1, \ldots, x_l, y_l \in G$ such that

$$\alpha(u)^{-1}\alpha(v) = [x_1, y_1] \cdots [x_l, y_l] = \prod_{i=1}^{l} [x_i, y_i]$$

(note that we do not have to consider the inverses of commutators since $[x, y]^{-1} = [y, x]$ in general). Since α is surjective, there exist $4l$ words $s_i, t_i, \bar{s}_i, \bar{t}_i \in A^*$ $(i = 1, \ldots, l)$ such that $\alpha(s_i) = x_i$, $\alpha(t_i) = y_i$, $\alpha(\bar{s}_i) = x_i^{-1}$, and $\alpha(\bar{t}_i) = y_i^{-1}$. Define two words $w, w' \in A^*$ by

$$w = \prod_{i=1}^{l} s_i t_i \bar{t}_i \bar{s}_i, \qquad w' = \prod_{i=1}^{l} s_i t_i \bar{s}_i \bar{t}_i.$$

Then we have

$$\alpha(w) = \prod_{i=1}^{l} x_i y_i y_i^{-1} x_i^{-1} = 1, \qquad \alpha(w') = \prod_{i=1}^{l} [x_i, y_i] = \alpha(u)^{-1}\alpha(v).$$

Note that w' is a rearrangement of w.

Consider two languages $\alpha^{-1}(\alpha(u))$ and A^*wA^* in REG_A. Their densities are $\delta_A(\alpha^{-1}(\alpha(u))) = \#(G)^{-1} > 0$ (by Lemma 2) and $\delta_A(A^*wA^*) = 1$ (by Example 1). The intersection $I = \alpha^{-1}(\alpha(u)) \cap A^*wA^*$ also has density $\delta_A(I) = \#(G)^{-1} > 0$ by Lemma 1.

To prove that L is Gcom-immeasurable, it suffices to show that every $K, M \in \mathsf{Gcom}_A$ with $K \subseteq L \subseteq M$ satisfies $I \subseteq M \setminus K$ since I has positive density. Let $s \in I$. Since $s \in \alpha^{-1}(\alpha(u))$ and $u \in L$, we have $s \in L \subseteq M$. Since $s \in A^*wA^*$, there exist $t_1, t_2 \in A^*$ such that $s = t_1 w t_2$. The rearrangement $s' = t_1 t_2 w'$ of s satisfies $\alpha(s') = \alpha(t_1)\alpha(w)\alpha(t_2)\alpha(w') = \alpha(s)\alpha(u)^{-1}\alpha(v) = \alpha(v)$, hence $s' \notin L \supseteq K$. Since $s' \notin K \in \mathsf{Gcom}_A$ and s' is a rearrangement of s, we have $s \notin K$. Thus $s \in M \setminus K$. $\square$

Corollary 1. *For any group language L whose syntactic monoid is non-commutative and for any Gcom-measurable language M, their symmetric difference $L \triangle M$ is an infinite set.*

Proof. Let $L \in \mathsf{G}_A$ and $M \in \mathrm{Ext}_A(\mathsf{Gcom}_A)$ such that M_L is non-commutative. Suppose contrarily that $L \triangle M$ is finite. Then $L \triangle M \in \mathrm{Ext}_A(\mathsf{Gcom}_A)$ and hence $(L \triangle M) \triangle M \in \mathrm{Ext}_A(\mathsf{Gcom}_A)$ since $\mathrm{Ext}_A(\mathsf{Gcom}_A)$ is closed under Boolean combination. Thus $L \in \mathrm{Ext}_A(\mathsf{Gcom}_A)$, which contradicts Theorem 1. □

4 Decidable characterisation of Gcom-measurability

The goal of this section is to prove the following Theorem 2, which gives the decidability of Gcom-measurability for regular languages. The following definition extends the Parikh mapping into the set of integer vectors so that we can use group theoretic tools. Intuitively, the set of vectors $\mathbb{Z}\mathrm{Pkh}_A(L) \cap \mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}})$ defined in Theorem 2 is a "boundary" between L and its complement from the viewpoint of commutative group languages. Hence, the equivalence of Condition (1) and Condition (2) in Theorem 2 states that a regular language L is Gcom-measurable if and only if the boundary $\mathbb{Z}\mathrm{Pkh}_A(L) \cap \mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}})$ has a *smaller rank* (than $\#(A)$). In other words, L is Gcom-measurable if and only if the boundary is *negligible* from the viewpoint of commutative group languages.

Definition 5 ($\mathbb{Z}\mathrm{Pkh}_A$). *Let $L \in \mathrm{REG}_A$ and choose a representation of the semilinear set $\mathrm{Pkh}_A(L) \subseteq \mathbb{N}^{\#(A)}$:*

$$\mathrm{Pkh}_A(L) = \bigcup_i (\boldsymbol{c}_i + \mathrm{span}_{\mathbb{N}}(\boldsymbol{p}_{i,1}, \ldots, \boldsymbol{p}_{i,\mu(i)})). \tag{1}$$

Then we define a semilinear set $\mathbb{Z}\mathrm{Pkh}_A(L) \subseteq \mathbb{Z}^{\#(A)}$ as

$$\mathbb{Z}\mathrm{Pkh}_A(L) \stackrel{\mathsf{def}}{=} \bigcup_i (\boldsymbol{c}_i + \mathrm{span}_{\mathbb{Z}}(\boldsymbol{p}_{i,1}, \ldots, \boldsymbol{p}_{i,\mu(i)})).$$

Note that $\mathbb{Z}\mathrm{Pkh}_A(L)$ depends on the choice (1) of representation of semilinear set. Our results are, however, true for any choice of representation.

Theorem 2. *Let $L \in \mathrm{REG}_A$ and*

$$\begin{aligned}
\mathbb{Z}\mathrm{Pkh}_A(L) &= \bigcup_i \Lambda_i, \quad \Lambda_i = \boldsymbol{c}_i + \mathrm{span}_{\mathbb{Z}}(\boldsymbol{p}_{i,1}, \ldots, \boldsymbol{p}_{i,\mu(i)}) \subseteq \mathbb{Z}^{\#(A)}, \\
\mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}}) &= \bigcup_j \Lambda'_j, \quad \Lambda'_j = \boldsymbol{c}'_j + \mathrm{span}_{\mathbb{Z}}(\boldsymbol{q}_{j,1}, \ldots, \boldsymbol{q}_{j,\nu(j)}) \subseteq \mathbb{Z}^{\#(A)}, \\
\mathbb{Z}\mathrm{Pkh}_A(L) \cap \mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}}) &= \bigcup_k \Lambda''_k, \quad \Lambda''_k = \boldsymbol{c}''_k + \mathrm{span}_{\mathbb{Z}}(\boldsymbol{r}_{k,1}, \ldots, \boldsymbol{r}_{k,\xi(k)}) \subseteq \mathbb{Z}^{\#(A)}.
\end{aligned}$$

Then the following are equivalent.

(1) L *is* Gcom*-measurable.*
(2) *For each* (i,j) *with* $\Lambda_i \cap \Lambda'_j \neq \varnothing$, $\operatorname{rank}(\Lambda_i) < \#(A)$ *or* $\operatorname{rank}(\Lambda'_j) < \#(A)$.
(3) *For each* k, $\operatorname{rank}(\Lambda''_k) < \#(A)$.
(4) *For each* (i,j) *with* $\Lambda_i \cap \Lambda'_j \neq \varnothing$,
   – $\dim(\operatorname{span}_{\mathbb{R}}(\boldsymbol{p}_{i,1}, \dots, \boldsymbol{p}_{i,\mu(i)})) < \#(A)$ *or*
   – $\dim(\operatorname{span}_{\mathbb{R}}(\boldsymbol{q}_{j,1}, \dots, \boldsymbol{q}_{j,\nu(j)})) < \#(A)$.
(5) *For each* k, $\dim(\operatorname{span}_{\mathbb{R}}(\boldsymbol{r}_{k,1}, \dots, \boldsymbol{r}_{k,\xi(k)})) < \#(A)$.

The intersection of two semilinear sets, which is again semilinear as stated in Section 2, and its rank can be effectively computable, thus we obtain the decidability of Gcom-measurability.

Corollary 2. *It is decidable whether a given regular language* L *is* Gcom*-measurable or not.*

Proving Theorem 2 involves several lemmata, propositions, and an approximation notion which we call *standard approximation.*

For each $n \in \mathbb{N}$, we write $\mathrm{Pkh}[n]_A \colon A^* \to (\mathbb{Z}/n\mathbb{Z})^{\#(A)}$ for the composition of $\mathrm{Pkh}_A \colon A^* \to \mathbb{N}^{\#(A)}$ and the natural surjection $\mathbb{N}^{\#(A)} \to (\mathbb{Z}/n\mathbb{Z})^{\#(A)}$.

Definition 6 (standard approximation). *Let* $L \subseteq A^*$ *be a (not necessarily regular) language. The* standard approximation *of* L *is the two sequences* $(\overline{L}_n)_{n\in\mathbb{N}}, (\underline{L}_n)_{n\in\mathbb{N}}$ *of regular languages defined as*

$$\overline{L}_n \overset{\mathsf{def}}{=\!=} \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L)), \qquad \underline{L}_n \overset{\mathsf{def}}{=\!=} \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))^{\mathrm{c}}.$$

It is easy to see that $\underline{L}_n \subseteq \underline{L}_{n+1} \subseteq L \subseteq \overline{L}_{n+1} \subseteq \overline{L}_n$ *for each* $n \in \mathbb{N}$*. That is,* $\overline{L}_n$ *(resp.* $\underline{L}_n$*) approximates* L from outer *(resp.* from inner*).*

Lemma 5. *For every* $L \in \mathsf{Gcom}_A$*, there exists some* $d \in \mathbb{N}$ *such that* L *is recognised by* $\mathrm{Pkh}[d]_A \colon A^* \to (\mathbb{Z}/d\mathbb{Z})^{\#(A)}$.

Proof. Let G be a finite commutative group and $\alpha \colon A^* \to G$ be a homomorphism recognising L. Define $d = \operatorname{lcm}\{\operatorname{ord}(\alpha(a)) \mid a \in A\}$, where $\operatorname{ord}(\alpha(a))$ denotes the *order* of $\alpha(a)$ in G. One can then define a well-defined group homomorphism $\varphi \colon (\mathbb{Z}/d\mathbb{Z})^A \to G$ such that $\varphi((x_a)_{a\in A}) = \prod_{a\in A} \alpha(a)^{x_a}$ for each $(x_a)_{a\in A} \in (\mathbb{Z}/d\mathbb{Z})^A$. It is easy to see that $\alpha = \varphi \circ \mathrm{Pkh}[d]_A$ and

$$L \subseteq \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(L)) \subseteq \alpha^{-1}(\alpha(L)) = L,$$

thus $\mathrm{Pkh}[d]_A$ recognises L. □

Lemma 6. *Let* $L \subseteq A^*$ *be a language and* $(\underline{L}_n \subseteq L \subseteq \overline{L}_n)_{n\in\mathbb{N}}$ *be the standard approximation of* L*. Then the following claims hold.*

(1) *If* $L \subseteq M \in \mathsf{Gcom}_A$*, then* $L \subseteq \overline{L}_n \subseteq M$ *for any sufficiently large* $n \in \mathbb{N}$.
(2) *If* $\mathsf{Gcom}_A \ni K \subseteq L$*, then* $K \subseteq \underline{L}_n \subseteq L$ *for any sufficiently large* $n \in \mathbb{N}$.

Proof.

(1) By Lemma 5, we may assume that M is recognised by $\mathrm{Pkh}[d]_A$ for some $d \in \mathbb{N}$. Let $n \in \mathbb{N}$ be sufficiently large so that d divides the factorial $n!$. Then the natural surjection $\rho_{n!} \colon (\mathbb{Z}/n!\mathbb{Z})^{\#(A)} \to (\mathbb{Z}/d\mathbb{Z})^{\#(A)}$ makes the diagram

$$\begin{tikzcd}
A^* \arrow[r, "\mathrm{Pkh}[n!]_A"] \arrow[dr, "\mathrm{Pkh}[d]_A"'] & (\mathbb{Z}/n!\mathbb{Z})^{\#(A)} \arrow[d, "\rho_{n!}"] \\
& (\mathbb{Z}/d\mathbb{Z})^{\#(A)}
\end{tikzcd}$$

commute. Thus

$$\begin{aligned}
L &\subseteq \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L)) = \overline{L}_n \\
&\subseteq \mathrm{Pkh}[n!]_A^{-1}(\rho_{n!}^{-1}(\rho_{n!}(\mathrm{Pkh}[n!]_A(L)))) \\
&= \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(L)) \subseteq \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(M)) = M.
\end{aligned}$$

(2) Similarly, we may assume that K is recognised by $\mathrm{Pkh}[d]_A$ and

$$\begin{aligned}
L = (L^{\mathrm{c}})^{\mathrm{c}} &\supseteq \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))^{\mathrm{c}} = \underline{L}_n \\
&\supseteq \mathrm{Pkh}[n!]_A^{-1}(\rho_{n!}^{-1}(\rho_{n!}(\mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))))^{\mathrm{c}} = \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(L^{\mathrm{c}}))^{\mathrm{c}} \\
&\supseteq \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(K^{\mathrm{c}}))^{\mathrm{c}}
\end{aligned}$$

since $\mathrm{Pkh}[d]_A$ is surjective and recognises K,

$$= \mathrm{Pkh}[d]_A^{-1}(\mathrm{Pkh}[d]_A(K)) = K. \square$$

Proposition 4. *For any language $L \subseteq A^*$ and its standard approximation $(\underline{L}_n \subseteq L \subseteq \overline{L}_n)_{n\in\mathbb{N}}$, the following are equivalent.*

(1) *L is Gcom-measurable.*
(2) $\lim_{n\to\infty} \delta_A(\overline{L}_n \setminus \underline{L}_n) = 0$.

Proof. The implication (2) $\implies$ (1) is obvious. Conversely, assume that L is Gcom-measurable. Then there exist sequences $(K_n \subseteq L \subseteq M_n)_{n\in\mathbb{N}}$ such that $K_n, M_n \in \mathsf{Gcom}_A$ and $\lim_{n\to\infty} \delta_A(M_n \setminus K_n) = 0$. By Lemma 6, there exists a non-decreasing sequence $(N(n))_{n\in\mathbb{N}}$ such that $K_n \subseteq \underline{L}_{N(n)} \subseteq L \subseteq \overline{L}_{N(n)} \subseteq M_n$ for each $n \in \mathbb{N}$. Thus $\lim_{n\to\infty} \delta_A(\overline{L}_n \setminus \underline{L}_n) = \lim_{n\to\infty} \delta_A(\overline{L}_{N(n)} \setminus \underline{L}_{N(n)}) \leq \lim_{n\to\infty} \delta_A(M_n \setminus K_n) = 0$. $\square$

Proposition 5. *For any language $L \subseteq A^*$ and its standard approximation $(\underline{L}_n \subseteq L \subseteq \overline{L}_n)_{n\in\mathbb{N}}$, we have*

$$\delta_A(\overline{L}_n \setminus \underline{L}_n) = \frac{\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))}{(n!)^{\#(A)}}.$$

Proof. We have

$$\begin{aligned}
\overline{L}_n \setminus \underline{L}_n &= \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L)) \setminus \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))^{\mathrm{c}} \\
&= \mathrm{Pkh}[n!]_A^{-1}(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))
\end{aligned}$$

and thus Lemma 2 completes the proof. $\square$

Proposition 6. *Let $L \in \mathrm{REG}_A$ and*

$$\mathbb{Z}\mathrm{Pkh}_A(L) = \bigcup_i \Lambda_i, \qquad \mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}}) = \bigcup_j \Lambda'_j,$$

where $\Lambda_i, \Lambda'_j \subseteq \mathbb{Z}^{\#(A)}$ are linear sets. Then the following are equivalent.

(1) $\displaystyle\lim_{n\to\infty} \frac{\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))}{(n!)^{\#(A)}} = 0.$

(2) *For any (i, j),*

$$\lim_{n\to\infty} \frac{\#\big(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j)\big)}{(n!)^{\#(A)}} = 0,$$

where $\pi_{n!}\colon \mathbb{Z}^{\#(A)} \to (\mathbb{Z}/n!\mathbb{Z})^{\#(A)}$ is the natural surjection.

Proof. Note that

$$\begin{aligned}\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}})) &= \#(\pi_{n!}(\mathbb{Z}\mathrm{Pkh}_A(L)) \cap \pi_{n!}(\mathbb{Z}\mathrm{Pkh}_A(L^{\mathrm{c}}))) \\ &= \#\left(\bigcup_{i,j}(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j))\right).\end{aligned}$$

(1) $\Longrightarrow$ (2). Contrarily suppose that

$$\lim_{n\to\infty} \frac{\#\big(\pi_{n!}(\Lambda_{i_0}) \cap \pi_{n!}(\Lambda'_{j_0})\big)}{(n!)^{\#(A)}} > 0$$

for some (i_0, j_0) (note that the limit exists because of the monotonicity). Then

$$\begin{aligned}\lim_{n\to\infty} \frac{\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))}{(n!)^{\#(A)}} &= \lim_{n\to\infty} \frac{\#\Big(\bigcup_{i,j}(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j))\Big)}{(n!)^{\#(A)}} \\ &\geq \lim_{n\to\infty} \frac{\#\big(\pi_{n!}(\Lambda_{i_0}) \cap \pi_{n!}(\Lambda'_{j_0})\big)}{(n!)^{\#(A)}} > 0.\end{aligned}$$

(2) $\Longrightarrow$ (1). We have

$$\begin{aligned}\lim_{n\to\infty} \frac{\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))}{(n!)^{\#(A)}} &= \lim_{n\to\infty} \frac{\#\Big(\bigcup_{i,j}(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j))\Big)}{(n!)^{\#(A)}} \\ \leq \lim_{n\to\infty} \frac{\sum_{i,j}\#\big(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j)\big)}{(n!)^{\#(A)}} &= \sum_{i,j} \lim_{n\to\infty} \frac{\#\big(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j)\big)}{(n!)^{\#(A)}} \\ = 0. \square\end{aligned}$$

Proposition 7. *Let $\Lambda_1, \Lambda_2 \subseteq \mathbb{Z}^{\#(A)}$ be two linear sets such that $\Lambda_1 \cap \Lambda_2 \neq \varnothing$. Then the following are equivalent.*

(1) $\lim_{n\to\infty} \dfrac{\#(\pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2))}{(n!)^{\#(A)}} = 0.$

(2) $\mathrm{rank}(\Lambda_1) < \#(A)$ *or* $\mathrm{rank}(\Lambda_2) < \#(A)$.

(3) $\mathrm{rank}(\Lambda_1 \cap \Lambda_2) < \#(A)$.

Proof.

(1) $\Longrightarrow$ (2). Contrarily suppose that $\mathrm{rank}(\Lambda_1) = \mathrm{rank}(\Lambda_2) = \#(A)$. Since the indices $\left(\mathbb{Z}^{\#(A)} : \Lambda_1\right)$ and $\left(\mathbb{Z}^{\#(A)} : \Lambda_2\right)$ are finite by Proposition 1, we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\#(\pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2))}{(n!)^{\#(A)}} = \lim_{n\to\infty} \frac{\#(\pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2))}{\#(\mathbb{Z}/n!\mathbb{Z})^{\#(A)}} \\
&= \lim_{n\to\infty} \frac{1}{\left((\mathbb{Z}/n!\mathbb{Z})^{\#(A)} : \pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2)\right)} \\
&\geq \lim_{n\to\infty} \frac{1}{\left((\mathbb{Z}/n!\mathbb{Z})^{\#(A)} : \pi_{n!}(\Lambda_1)\right)} \cdot \frac{1}{\left((\mathbb{Z}/n!\mathbb{Z})^{\#(A)} : \pi_{n!}(\Lambda_2)\right)} \\
&= \lim_{n\to\infty} \frac{1}{\left(\mathbb{Z}^{\#(A)} : \Lambda_1 + n!\mathbb{Z}^{\#(A)}\right)} \cdot \frac{1}{\left(\mathbb{Z}^{\#(A)} : \Lambda_2 + n!\mathbb{Z}^{\#(A)}\right)} \\
&\geq \frac{1}{\left(\mathbb{Z}^{\#(A)} : \Lambda_1\right)} \cdot \frac{1}{\left(\mathbb{Z}^{\#(A)} : \Lambda_2\right)} > 0.
\end{aligned}$$

(2) $\Longrightarrow$ (1). We may assume that $\mathrm{rank}(\Lambda_1) < \#(A)$. Then we have

$$\begin{aligned}
&\lim_{n\to\infty} \frac{\#(\pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2))}{(n!)^{\#(A)}} = \lim_{n\to\infty} \frac{1}{\left((\mathbb{Z}/n!\mathbb{Z})^{\#(A)} : \pi_{n!}(\Lambda_1) \cap \pi_{n!}(\Lambda_2)\right)} \\
&\leq \lim_{n\to\infty} \frac{1}{\left((\mathbb{Z}/n!\mathbb{Z})^{\#(A)} : \pi_{n!}(\Lambda_1)\right)} = 0.
\end{aligned}$$

The last equality is exactly Proposition 2.

(2) $\Longrightarrow$ (3). $\mathrm{rank}(\Lambda_1 \cap \Lambda_2) \leq \min\{\mathrm{rank}(\Lambda_1), \mathrm{rank}(\Lambda_2)\} < \#(A)$.

(3) $\Longrightarrow$ (2). Contrarily suppose that $\mathrm{rank}(\Lambda_1) = \mathrm{rank}(\Lambda_2) = \#(A)$. Since Λ_1, Λ_2 are coset of $\mathbb{Z}^{\#(A)}$, there exist constant vectors $\boldsymbol{c}_1, \boldsymbol{c}_2 \in \mathbb{Z}^{\#(A)}$ and subgroups $M_1, M_2 \subseteq \mathbb{Z}^{\#(A)}$ such that $\Lambda_1 = \boldsymbol{c}_1 + M_1, \Lambda_2 = \boldsymbol{c}_2 + M_2$. Since $\Lambda_1 \cap \Lambda_2 \neq \varnothing$, there exists $\boldsymbol{c} \in M_1 \cap M_2$ such that $\Lambda_1 \cap \Lambda_2 = \boldsymbol{c} + (M_1 \cap M_2)$. Since $\mathrm{rank}(M_1) = \mathrm{rank}(\Lambda_1) = \#(A)$ and $\mathrm{rank}(M_2) = \mathrm{rank}(\Lambda_2) = \#(A)$, by Proposition 1, there exists $N_1, N_2 \in \mathbb{N}$ such that $N_1\mathbb{Z}^{\#(A)} \subseteq M_1$ and $N_2\mathbb{Z}^{\#(A)} \subseteq M_2$. Thus $N_1N_2\mathbb{Z}^{\#(A)} \subseteq M_1 \cap M_2$ and $\#(A) \geq \mathrm{rank}(\Lambda_1 \cap \Lambda_2) = \mathrm{rank}(M_1 \cap M_2) \geq \mathrm{rank}(N_1N_2\mathbb{Z}^{\#(A)}) = \#(A)$. □

Proof (of Theorem 2). The equivalences (2) $\iff$ (4) and (3) $\iff$ (5) are clear. For the standard approximation $(\underline{L}_n \subseteq L \subseteq \overline{L}_n)_{n\in\mathbb{N}}$ of L,

$$\begin{aligned}
(1) \iff & \lim_{n\to\infty} \delta_A(\overline{L}_n \setminus \underline{L}_n) = 0 && \text{(Proposition 4)}\\
\iff & \lim_{n\to\infty} \frac{\#(\mathrm{Pkh}[n!]_A(L) \cap \mathrm{Pkh}[n!]_A(L^{\mathrm{c}}))}{(n!)^{\#(A)}} = 0 && \text{(Proposition 5)}\\
\iff & \forall i,j \left[\lim_{n\to\infty} \frac{\#\big(\pi_{n!}(\Lambda_i) \cap \pi_{n!}(\Lambda'_j)\big)}{(n!)^{\#(A)}} = 0 \right] && \text{(Proposition 6)}\\
\iff & \forall i,j \left[\begin{matrix} \Lambda_i \cap \Lambda'_j = \varnothing \vee \\ \mathrm{rank}(\Lambda_i) < \#(A) \vee \mathrm{rank}(\Lambda'_j) < \#(A) \end{matrix} \right] && \text{(Proposition 7)}\\
& (\iff (2))\\
\iff & \forall k[\mathrm{rank}(\Lambda''_k) < \#(A)] && \text{(Proposition 7)}\\
& (\iff (3)).\square
\end{aligned}$$

5 Simple characterisation of MOD-measurability

The decidability of MOD-measurability for regular languages is essentially already given by Theorem 2. Moreover, we have a simple language theoretic characterisation of MOD-measurable languages as follows.

Theorem 3. *For any $L \in \mathrm{REG}_A$, the following are equivalent.*

(1) *L is* MOD*-measurable.*
(2) *$\mathrm{len}_A(L) \cap \mathrm{len}_A(L^{\mathrm{c}}) \subseteq \mathbb{N}$ is a finite set, where $\mathrm{len}_A \colon A^* \to \mathbb{N}$ is the length function $w \mapsto |w|$.*
(3) *There exists a unique $M \in \mathsf{MOD}_A$ such that $L \triangle M$ is a finite set.*

Proof. The proof of the equivalence (1) $\iff$ (2) is similar to that of Theorem 2. Indeed, it suffices to replace all the occurrences of Gcom, Pkh, and $\#(A)$ in Section 4 by MOD, len, and 1, respectively. (Note that $\mathbb{Z}\mathrm{len}_A(L) \cap \mathbb{Z}\mathrm{len}_A(L^{\mathrm{c}}) \subseteq \mathbb{Z}$ is finite if and only if $\mathrm{len}_A(L) \cap \mathrm{len}_A(L^{\mathrm{c}}) \subseteq \mathbb{N}$ is finite.)

(3) $\Longrightarrow$ (1). From the assumption, $L \triangle M$ and M are both MOD-measurable. The class $\mathrm{Ext}_A(\mathsf{MOD}_A)$ is closed under Boolean combination, hence $L = (L \triangle M) \triangle M \in \mathrm{Ext}_A(\mathsf{MOD}_A)$.

(2) $\Longrightarrow$ (3). Since $\mathrm{len}_A(L)$ is a semilinear set of $\mathbb{N}$, we may assume that

$$\mathrm{len}_A(L) = \bigcup_i (c_i + \mathrm{span}_{\mathbb{N}}(p_i)) \cup \bigcup_j \{d_j\} \quad (p_i > 0 \text{ for each } i).$$

Define $S = \bigcup_i (c'_i + \mathrm{span}_{\mathbb{N}}(p_i))$, where $c'_i = c_i \bmod p_i$. Then we have $\mathrm{len}_A^{-1}(S) \in \mathsf{MOD}_A$. Since $\mathrm{len}_A(L) \triangle S$ is a finite set by construction, the inverse image $\mathrm{len}_A^{-1}(\mathrm{len}_A(L) \triangle S) = \mathrm{len}_A^{-1}(\mathrm{len}_A(L)) \triangle \mathrm{len}_A^{-1}(S)$ is also finite. Since we have

$$L \triangle \mathrm{len}_A^{-1}(\mathrm{len}_A(L)) = \mathrm{len}_A^{-1}(\mathrm{len}_A(L)) \cap L^{\mathrm{c}} \subseteq \mathrm{len}_A^{-1}(\mathrm{len}_A(L) \cap \mathrm{len}_A(L^{\mathrm{c}})),$$

the set $L \triangle \mathrm{len}_A^{-1}(\mathrm{len}_A(L))$ is finite by the assumption (2). Thus

$$L \triangle \mathrm{len}_A^{-1}(S) = (L \triangle \mathrm{len}_A^{-1}(\mathrm{len}_A(L))) \triangle (\mathrm{len}_A^{-1}(\mathrm{len}_A(L)) \triangle \mathrm{len}_A^{-1}(S))$$

is a finite set.

Suppose that there exist two languages $M_1, M_2 \in \mathsf{MOD}_A$ such that both $L \triangle M_1$ and $L \triangle M_2$ are finite sets. Then $M_1 \triangle M_2 = (L \triangle M_1) \triangle (L \triangle M_2)$ is also a finite set. Since MOD_A is closed under Boolean combination, $M_1 \triangle M_2 \in \mathsf{MOD}_A$. Thus $M_1 = M_2$ since the only finite set in MOD_A is $\varnothing$. $\square$

6 Conclusion and future work

As we described in Section 1, while the measuring power of subclasses of star-free languages are systematically studied so far, nothing is known for the measuring power of group languages before. This paper gave the first decidability results on this topic: the Gcom-measurability and the MOD-measurability for regular languages are both decidable thanks to Theorem 2. Also, Theorem 1 tells us that there is a huge gap between group and commutative group languages even from a (very rough) measure theoretic point of view. To clarify the computational complexity of these two measurability and the decidability of the G-measurability for regular languages are our important future work. Also, to give a purely algebraic characterisation of the G-measurability (the Gcom-measurability, respectively) is an interesting open problem for us. For the case of MOD-measurability, we have very simple language theoretic characterisation as stated in Theorem 3. We are interested whether it is possible to obtain a similar language theoretic characterisation of Gcom-measurability.

No different characterisation of REG-measurability is yet known, and only few examples of REG-*im*measurable languages are known (*cf.* [6]). The Krohn-Rhodes theorem states that, for every finite monoid M, there exists a sequence $G_1, \ldots, G_n$ of finite groups dividing M and a sequence $M_0, \ldots, M_n$ of aperiodic finite monoids such that M divides $M_0 \circ G_1 \circ M_1 \circ \cdots \circ G_n \circ M_n$ where $\circ$ is the wreath product operation (*cf.* [3]). Roughly speaking, this means that every regular language can be represented as some "combination" of group and star-free languages because a language is star-free if and only if its syntactic monoid is aperiodic. In this sense, we can consider the class of group languages and the class of star-free languages as two representative subclasses of regular languages. We hope that a deep understanding of the REG-measurable languages (or, group languages its self) might be obtained by a further study of the measuring power of group languages.

Acknowledgements: The authors thank to anonymous reviewers for many valuable comments. This work was supported by JSPS KAKENHI Grant Number JP20H05961 and JST ACT-X Grant Number JPMJAX210B.

References

1. Berstel, J., Perrin, D., Reutenauer, C.: Codes and Automata, Encyclopedia of mathematics and its applications, vol. 129. Cambridge University Press (2010)
2. Lawson, M.V.: Finite Automata. Chapman and Hall/CRC (2004)
3. Pin, J.E.: Mathematical foundations of automata theory. `https://www.irif.fr/~jep/PDF/MPRI/MPRI.pdf`, version of Feburary 18, 2022
4. Place, T., Zeitoun, M.: Group separation strikes back. In: 2023 38th Annual ACM/IEEE Symposium on Logic in Computer Science (LICS). pp. 1–13. IEEE Computer Society, Los Alamitos, CA, USA (jun 2023)
5. Shallit, J.: A Second Course in Formal Languages and Automata Theory. Cambridge University Press, USA, 1 edn. (2008)
6. Sin'ya, R.: Asymptotic approximation by regular languages. In: Current Trends in Theory and Practice of Computer Science. pp. 74–88 (2021)
7. Sin'ya, R.: Carathéodory extensions of subclasses of regular languages. In: Developments in Language Theory. pp. 355–367 (2021)
8. Sin'ya, R.: Measuring power of locally testable languages. In: Diekert, V., Volkov, M. (eds.) Developments in Language Theory. pp. 274–285. Springer International Publishing, Cham (2022)
9. Sin'ya, R.: Measuring power of generalised definite languages. In: Implementation and Application of Automata. pp. 278–289. Springer International Publishing (2023)
10. Sin'ya, R., Yamaguchi, Y., Nakamura, Y.: Regular languages that can be approximated by testing subword occurrences. Computer Software **40**(2), 49–60 (2023), (written in Japanese)
11. Thierrin, G.: Permutation automata. Mathematical systems theory **2**, 83–90 (1968)

A Appendix

Proposition 8 (Smith normal form). *Let $s < r$ and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s \in \mathbb{Z}^r$ be r-dimensional column vectors, and $X = \begin{bmatrix}\boldsymbol{p}_1 \cdots \boldsymbol{p}_s\end{bmatrix} \in \mathbb{Z}^{r\times s}$ be an $r \times s$ matrix. Then there exists a pair (P, Q) of unimodular matrices $P \in \mathbb{Z}^{r\times r}$ and $Q \in \mathbb{Z}^{s\times s}$ such that*

$$PXQ = \left[\begin{array}{c} \begin{matrix} e_1 & & \mathbf{0} \\ & e_2 & \\ \mathbf{0} & & \ddots \\ & & & e_s \end{matrix} \\ \hline \mathbf{0} \end{array}\right], \qquad e_1 \mid e_2 \mid \cdots \mid e_s,$$

where $x \mid y$ mean that x divides y. Note that any $e \in \mathbb{N}$ divides 0, i.e., $e \mid 0$. Moreover, there exists an isomorphism $\mathbb{Z}^r / \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s) \cong \mathbb{Z}/e_1\mathbb{Z} \times \mathbb{Z}/e_2\mathbb{Z} \times \cdots \times \mathbb{Z}/e_s\mathbb{Z}$. In particular, if $e_s \neq 0$, then $(\mathbb{Z}^r : \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s)) = e_1 e_2 \cdots e_s$. □

Example 3. Let $X = \left[\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right]$. Not that X is not in Smith normal form since $2 \nmid 3$. Then, by applying elementary row and column operations, we have

$$\begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & 0 \\ 3 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 2 & 0 \\ -1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 3 \\ -1 & 3 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 \\ -1 & 6 \end{bmatrix} \rightsquigarrow \begin{bmatrix} 1 & 0 \\ 0 & 6 \end{bmatrix}.$$

Thus $\mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/3\mathbb{Z} \cong \mathbb{Z}^2 / \operatorname{span}_{\mathbb{Z}}(\left[\begin{smallmatrix} 2 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 3 \end{smallmatrix}\right]) \cong \mathbb{Z}^2 / \operatorname{span}_{\mathbb{Z}}(\left[\begin{smallmatrix} 1 \\ 0 \end{smallmatrix}\right], \left[\begin{smallmatrix} 0 \\ 6 \end{smallmatrix}\right]) \cong \mathbb{Z}/6\mathbb{Z}$.

Proof (Proof of Proposition 2). Let $\operatorname{rank}(M) = s < r$ and $\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s \in \mathbb{Z}^r$ be a free basis of M. Define an $r \times s$ matrix $X \in \mathbb{Z}^{r\times s}$ as $X = \begin{bmatrix}\boldsymbol{p}_1 \cdots \boldsymbol{p}_s\end{bmatrix}$. Then, by Proposition 8, there exist two unimodular matrices $P \in \mathbb{Z}^{r\times r}$ and $Q \in \mathbb{Z}^{s\times s}$ such that

$$PXQ = \left[\begin{array}{c} \begin{matrix} e_1 & & \mathbf{0} \\ & e_2 & \\ \mathbf{0} & & \ddots \\ & & & e_s \end{matrix} \\ \hline \mathbf{0} \end{array}\right], \qquad e_1 \mid e_2 \mid \cdots \mid e_s.$$

Here $e_1, e_2, \ldots, e_s$ are all non-zero since $\operatorname{rank}(M) = s$. Let $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_r$ be the standard basis of $\mathbb{Z}^r$. Define an $r \times (s + r)$ matrix $X(t) \in \mathbb{Z}[t]^{r\times(s+r)}$ as $X(t) = \begin{bmatrix}\boldsymbol{p}_1 \cdots \boldsymbol{p}_s \; t\boldsymbol{e}_1 \cdots t\boldsymbol{e}_r\end{bmatrix}$. Then there exist two matrices $Y \in \mathbb{Z}^{s\times r}$ and $Z \in \mathbb{Z}^{(r-s)\times r}$ such that

$$P \cdot X(t) \cdot \begin{bmatrix} Q & O \\ O & I_r \end{bmatrix} = \left[\begin{array}{c|c} \begin{matrix} e_1 & & \mathbf{0} \\ & e_2 & \\ \mathbf{0} & & \ddots \\ & & & e_s \end{matrix} & tY \\ \hline \mathbf{0} & tZ \end{array}\right],$$

where $I_r \in \mathbb{Z}^{r\times r}$ is the identity matrix. Suppose that the Smith normal form of Z is

$$\left[\begin{array}{cccc|c} e_{s+1} & & & \mathbf{0} & \\ & e_{s+2} & & & \\ & & \ddots & & \mathbf{0} \\ \mathbf{0} & & & e_r & \end{array}\right].$$

If $n \geq \max\{e_1, e_2, \ldots, e_s\}$, then the Smith normal form of $X(n!)$ is

$$\left[\begin{array}{cccc|cccc|c} e_1 & & & \mathbf{0} & & & & & \\ & e_2 & & & & \mathbf{0} & & & \mathbf{0} \\ & & \ddots & & & & & & \\ \mathbf{0} & & & e_s & & & & & \\ \hline & & & & n!e_{s+1} & & & \mathbf{0} & \\ & \mathbf{0} & & & & n!e_{s+2} & & & \mathbf{0} \\ & & & & & & \ddots & & \\ & & & & \mathbf{0} & & & n!e_r & \end{array}\right].$$

Here $e_{s+1}, e_{s+2}, \ldots, e_r$ are all non-zero since $\operatorname{rank}(M + n!\mathbb{Z}^r) = r$. Hence we have

$$\begin{aligned}
&\lim_{n\to\infty} ((\mathbb{Z}/n!\mathbb{Z})^r : \pi_{n!}(M)) \\
&= \lim_{n\to\infty} (\mathbb{Z}^r : M + \operatorname{Ker}(\pi_{n!})) = \lim_{n\to\infty} (\mathbb{Z}^r : M + n!\mathbb{Z}^r) \\
&= \lim_{n\to\infty} (\mathbb{Z}^r : \operatorname{span}_{\mathbb{Z}}(\boldsymbol{p}_1, \ldots, \boldsymbol{p}_s, n!\boldsymbol{e}_1, \ldots, n!\boldsymbol{e}_r)) \\
&= \lim_{n\to\infty} \# \begin{pmatrix} \mathbb{Z}/e_1\mathbb{Z} \times \mathbb{Z}/e_2\mathbb{Z} \times \cdots \times \mathbb{Z}/e_s\mathbb{Z} \times \\ \mathbb{Z}/n!e_{s+1}\mathbb{Z} \times \mathbb{Z}/n!e_{s+2}\mathbb{Z} \times \cdots \times \mathbb{Z}/n!e_r\mathbb{Z} \end{pmatrix} \\
&= \lim_{n\to\infty} e_1 e_2 \cdots e_s (n!e_{s+1})(n!e_{s+2}) \cdots (n!e_r) = \infty. \square
\end{aligned}$$