Measuring Power of Locally Testable Languages

Ryoma Sin'ya (Akita University, Japan)

DLT 2022 May 12 @USF Tampa, US.

秋田大学
Akita University

Outline

1. Background I: measurability (5 min.)
2. Background II: fragments of star-free languages (5 min.)
3. Measuring power of locally testable languages (10 min.)
4. Conclusion (2~5 min.)

Outline

1. Background I: measurability
2. Background II: fragments of star-free languages
3. Measuring power of locally testable languages
4. Conclusion

Density of formal languages

The density of a language L over A is defined as

$$\delta_A(L) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=0}^{n-1} \frac{\#(L \cap A^i)}{\#(A^i)}.$$

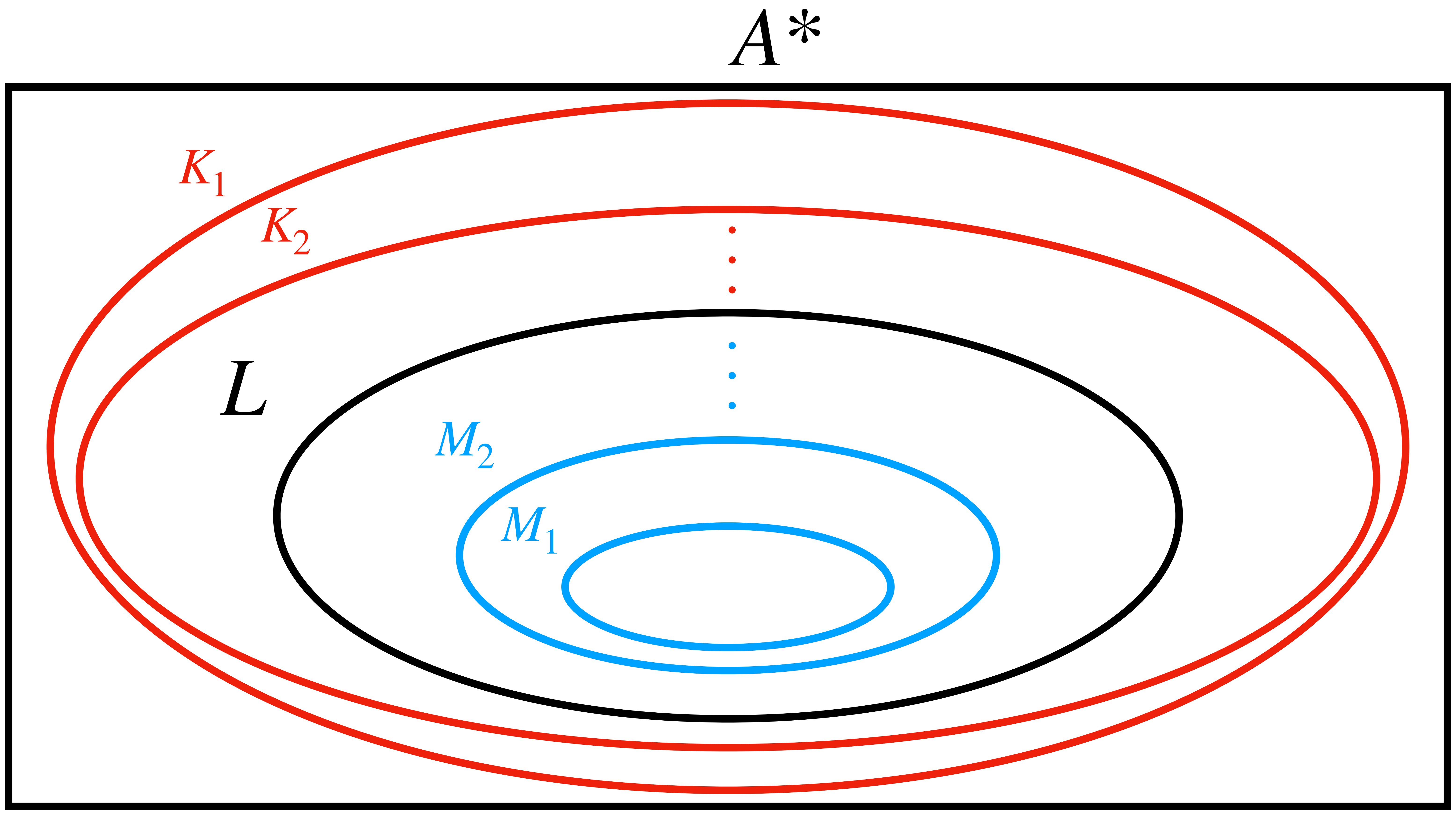
Example 1: $\delta_A((AA)^*) = \dfrac{1}{2}$.

Example 2: $\delta_A(aA^*) = \dfrac{1}{\#(A)}$.

Example 3: $L_\perp = \{w \in A^* \mid 3^n \le |w| < 3^{n+1}$ for some even $n\}$ does **not** have a density.

Theorem (cf. [Berstel 1973]):
Every regular language *do have a rational density*.

$\mathscr{C}$-measurability [S., SOFSEM'21] (cf. [Buck, 1946])



L is said to be $\mathscr{C}$-**measurable** if there exists an *infinite sequence of pairs of languages* $(M_n, K_n)_{n\in\mathbb{N}}$ in $\mathscr{C}$ such that $M_n \subseteq L \subseteq K_n$ and $\lim_{n\to\infty} \delta_A(K_n \setminus M_n) = 0$.

Example of a regular measurable language

Theorem [S., SOFSEM'21]:
D $= \{\varepsilon, ab, aabb, abab, \dots\}$ over $A = \{a, b\}$ is regular measurable.

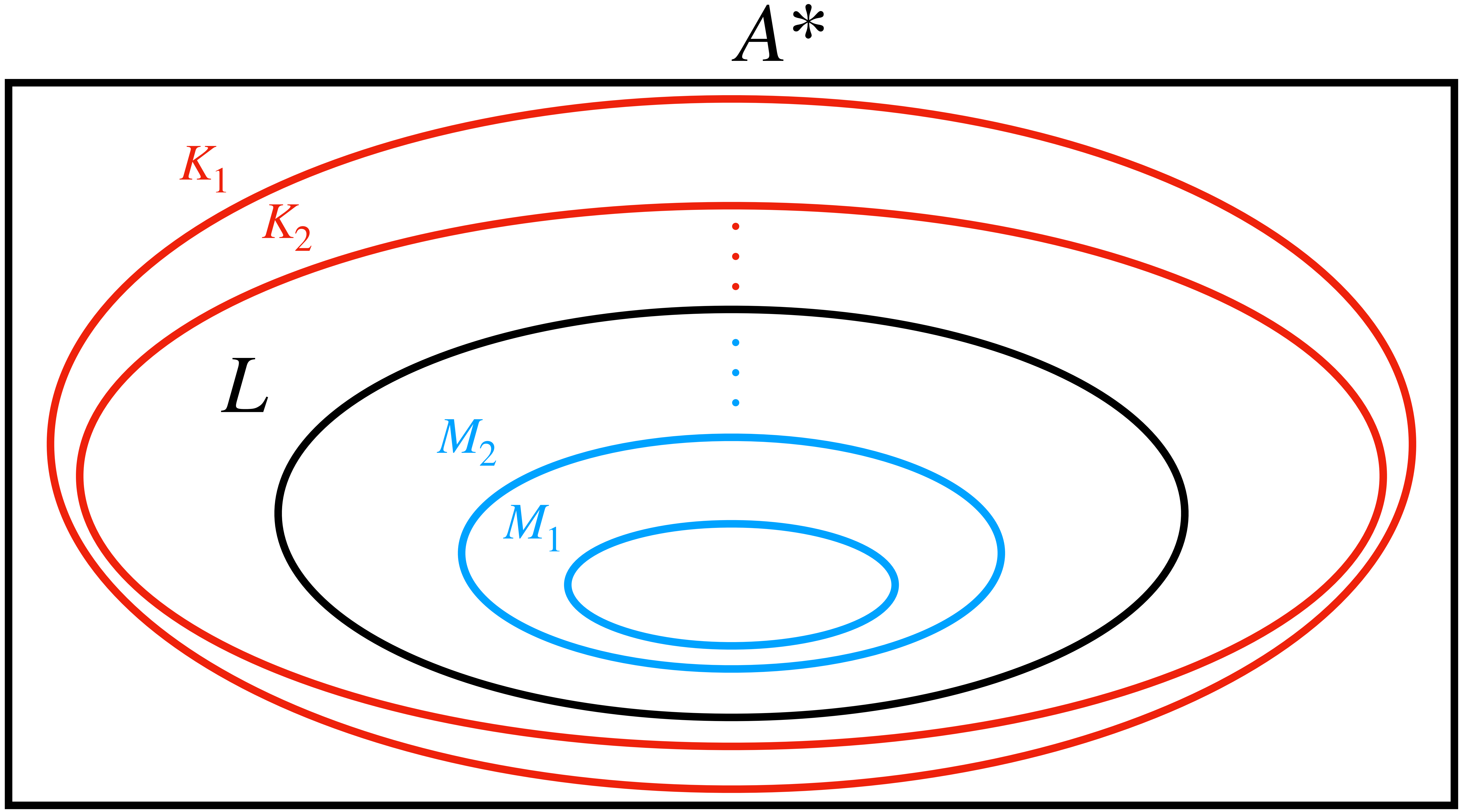
Proof: Let $L_k = \{w \in A^* \mid |w|_a = |w|_b \mod k\}$ for each $k \geq 1$.

the # of occurrences of a in w

Then, for each $k \geq 1$, D $\subseteq L_k$ and $\delta_A(L_k) = \dfrac{1}{k} \to 0$ (if $k \to \infty$).

Thus the infinite sequence $(\emptyset, L_k)_{k \geq 1}$ converges to D.

$\mathscr{C}$-measurability [S., SOFSEM'21] (cf. [Buck, 1946])



L is said to be $\mathscr{C}$**-measurable** if there exists an *infinite sequence of pairs of languages* $(M_n, K_n)_{n\in\mathbb{N}}$ in $\mathscr{C}$ such that $M_n \subseteq L \subseteq K_n$ and $\lim_{n\to\infty} \delta_A(K_n \backslash M_n) = 0$.

Original motivation of mesurability

- A non-empty word w is said to be ***primitive*** if it can not be represented as a power of shorter words, i.e., $w = u^n \Rightarrow u = w$ (and $n = 1$).
  Q denotes the set of all primitive words over $\{a, b\}$.

Example： $ababa \in$ Q $ababab = (ab)^3 \notin$ Q

Conjecture [Dömösi-Horvath-Ito 1991]: Q is not context-free.

Regular measurability was originally introduced for tackling this conjecture (cf. [S., SOFSEM’21] [S., DLT’21]).


[S., DLT'21]
SF = "Star Free"
"Unambiguous Polynomial" = UPol
LT = "Locally Testable"
"Piecewise Testable" = PT
This work [S.,DLT'22]
AT = "Alphabet Testable"

Outline

1. Background I: measurability
2. Background II: fragments of star-free languages
3. Measuring power of locally testable languages
4. Conclusion


SF = "Star Free"
"Unambiguous Polynomial" = UPol
LT = "Locally Testable"
"Piecewise Testable" = PT
AT = "Alphabet Testable"

Testable languages

A language L over A is called

- ***locally testable*** if it can be represented as a finite Boolean combination of languages of the form uA^*, A^*v and A^*wA^* [McNaughton-Papert 1971].

- ***piecewise testable*** if it can be represented as a finite Boolean combination of languages of the form $L_w = A^*a_1A^*a_2\ldots A^*a_nA^*$ where $w = a_1a_2\cdots a_n$ [Simon 1972].

- ***alphabet testable*** if it can be represented as a finite Boolean combination of languages of the form A^*aA^* (where $a \in A$) (cf. [Place-Zeitoun 2021]).


SF = "Star Free"
"Unambiguous Polynomial" = UPol
LT = "Locally Testable"
"Piecewise Testable" = PT
AT = "Alphabet Testable"

Unambiguous polynomials [Schützenberger 1976]

- **A *monomial*** is a language of the form $A_0^* a_1 A_1^* \ldots A_{n-1}^* a_n A_n^*$ where $a_i \in A$ *and* $A_i \subseteq A$.

- A monomial $M = A_0^* a_1 A_1^* \ldots A_{n-1}^* a_n A_n^*$ is ***unambiguous*** if it has the unique factorisation property *i.e.,* $\forall w \in M, \ \exists ! w_0, \ldots, w_n \in A^*$ s.t.
$$w = w_0 a_1 w_1 \cdots a_n w_n \quad \text{and} \quad w_i \in A_i^* \text{ for each } i.$$

- A language is called ***unambiguous polynomial*** if it can be represented as a finite disjoint union of unambiguous monomials.


$\emptyset^* a \emptyset^* b \emptyset^* c A^* =$
$abcA^*$
$\Subset$
SF = “Star Free”
$abcA^*$
$\Subset$
“Unambiguous Polynomial” = UPol
$\not\Supset$
A^*abcA^*
$\in$
LT = “Locally Testable”
$\Supset$
$\notin$
$A^*aA^*bA^*cA^*$
$= (A\backslash\{a\})^*a(A\backslash\{b\})^*b(A^*\backslash\{c\})cA^*$
“Piecewise Testable” = PT
$\not\Cup$
$abcA^*$
AT = “Alphabet Testable”

A Survey on Small Fragments of First-Order Logic over Finite Words

Volker Diekert[1], Paul Gastin[2], Manfred Kufleitner[1,3]

[1] FMI, Universität Stuttgart
Universitätsstr. 38, D-70569 Stuttgart, Germany
{diekert,kufleitner}@fmi.uni-stuttgart.de

[2] LSV, ENS Cachan, CNRS
61, Av. du Président Wilson, F-94230 Cachan, France
Paul.Gastin@lsv.ens-cachan.fr

[3] LaBRI, Université de Bordeaux & CNRS
351, Cours de la Libération, F-33405 Talence Cedex, France

Outline

1. Background I: measurability
2. Background II: fragments of star-free languages
3. Measuring power of locally testable languages
4. Conclusion

Main results

Theorem I:
L is LT-measurable if and only if L is UPol-measurable .

↓Easy

(relatively) Not easy↑

Theorem II:
L is PT-measurable if and only if L or its complement contains L_w for some w.

Theorem III:
L is AT-measurable if and only if L or its complement contains $\bigcap_{a \in A} A^*aA^*$.

Theorem I:
L is LT-measurable if and only if L is UPol-measurable.

Proof sketch of "L is LT-measurable $\Rightarrow$ L is UPol-measurable".

Remark: it is enough to show that any language of the form uA^*, A^*v, A^*wA^* are UPol-measurable (since UPol-measurability is closed under Boolean operations).

$uA^* = \emptyset^* a_1 \emptyset^* \cdots \emptyset^* a_n A^*$ itself is an unambiguous polynomial where $u = a_1 \cdots a_n$, hence uA^* is UPol-measurable (A^*v is UPol-measurable, too).

However, A^*wA^*, is not a unambiguous polynomial in general (like A^*abcA^*). We should construct a convergent sequence of unambiguous polynomials.

Theorem I:
L is LT-measurable if and only if L is UPol-measurable.

Proof sketch of "L is LT-measurable $\Rightarrow$ L is UPol-measurable".

$$A^*wA^* = \biguplus_{n\geq 0}^{\infty} \left\{ x \in A^* \mid w \text{ firstly apperas in } x \text{ at the index } n \text{ as a factor} \right\}.$$

Each language $W_n = \left\{ x \in A^* \mid w \text{ firstly apperas in } x \text{ at the index } n \text{ as a factor} \right\}$ is an unambiguous polynomial, hence its finite disjoint union $\biguplus_{n\geq 0}^{k} W_k$ converges to A^*wA^* from inner.

Theorem I:
L is LT-measurable if and only if L is UPol-measurable.

(very rough) Proof outline of "L is LT-measurable $\Leftarrow$ L is UPol-measurable".

Let $M = A_0^* a_1 A_1^* \cdots a_n A_n^*$ be an unambiguous monomial.

We construct a convergent sequence $(L_n)_{n \in \mathbb{N}}$ of locally testable languages from **inner**, *i.e.*, $L_n \subseteq M$ and $\lim_{n \to \infty} \delta_A(L_n) = \delta_A(M)$.

If we can construct such a sequence, then we can also construct a convergent sequence of locally testable languages from **outer** (because the complement of an unambiguous polynomial is also unambiguous polynomial [Schützenberger 1976]).

Theorem I:
L is LT-measurable if and only if L is UPol-measurable.

(very rough) Proof outline of "L is LT-measurable $\Leftarrow$ L is UPol-measurable".

Let $M = A_0^* a_1 A_1^* \cdots a_n A_n^*$ be an unambiguous monomial.

We construct a convergent sequence $(L_n)_{n \in \mathbb{N}}$ of locally testable languages from **inner**, *i.e.*, $L_n \subseteq M$ and $\lim_{n \to \infty} \delta_A(L_n) = \delta_A(M)$.

Case $\delta_A(M) = 0$: The constant sequence $(\varnothing)_{n \in \mathbb{N}}$ satisfies the condition.

Theorem I:
L is LT-measurable if and only if L is UPol-measurable.

(very rough) Proof outline of "L is LT-measurable $\Leftarrow$ L is UPol-measurable".

Case $\delta_A(M) > 0$:

Lemma 1: M can be written as $M = PA^*S$ where $\delta_A(P) = \delta_A(S) = 0$.

Lemma 2: $PA^*S = \biguplus_{n\geq 0} \{xA^*z \mid (x,z) \in U_n\}$ where U_n is some finite set.

The sequence $\left(\bigcup_{(x,z)\in U_n} xA^*z \right)_{n\in\mathbb{N}}$ of locally testable languages converges to M.

Theorem II:
L is PT-measurable if and only if L or its complement contains L_w for some w.

Theorem II' (algebraic characterisation of PT-measurable regular languages):
A regular language L is PT-measurable if and only if its syntactic monoid M and syntactic morphism η satisfies the following condition: for every $x \in M \backslash \{0\}$ there is a letter $a \in A$ such that $x'\eta(a) <_{\mathcal{J}} x'$ for every $x' \mathcal{J} x$.

Corollary: the PT-measurability for regular languages is decidable.

Outline

1. Background I: density and measurability
2. Background II: fragments of star-free languages
3. Measuring power of locally testable languages
4. Conclusion

Main results

Theorem I:
L is LT-measurable if and only if L is UPol-measurable .

Theorem II:
L is PT-measurable if and only if L or its complement contains L_w for some w.

Theorem III:
L is AT-measurable if and only if L or its complement contains $\bigcap_{a \in A} A^*aA^*$.

Decision problems and complexity (progress report)

- Yoshiki Nakamura pointed out that the PT-measurability for regular languages is **decidable in linear time**, by using a reduction to some decision problem considered in [N. Rampersad, J. Shallit, Z. Xu, 2009].
- Yutaro Yamaguchi pointed out that the AT-measurability for regular languages is **coNP-complete**.
  - The same coNP-completeness result was independently shown in Kazuhiro Inaba's unpublished manuscript
    "Quick Brown Fox in Formal Languages"
    on arXiv uploaded in 2015.

- We summarised these results in our manuscript written in Japanese [S.-Yamaguchi-Nakamura, PPL2022].

Open problems and future work

- Is the LT-measurability for regular languages decidable?

- Does “L is SF-measurable $\Leftrightarrow$ L is LT-measurable” hold or not?

- Measuring power of other fragments (or super class) of regular languages?

Thanks!

(Akita-Inu)

秋田大学
Akita University

Summary (“Table 1” of my paper)

Language	**Algebra**	**Logic**	**Measurability**
SF	aperiodic	FO	$\mathrm{SF} \subsetneq \mathrm{RExt}_A(\mathrm{SF}) \subsetneq \mathrm{REG}$ [22]
LT	locally idempotent and commutative		$\mathrm{Ext}_A(\mathrm{LT}) = \mathrm{Ext}_A(\mathrm{UPol})$
UPol	**DA**	FO^2	
PT	$\mathcal{J}$-trivial	$\mathbb{B}\Sigma_1$	$\mathrm{PT} \subsetneq \mathrm{RExt}_A(\mathrm{PT}) \subsetneq \mathrm{ZO}$ L is PT-measurable iff L or $\overline{L}$ contains a simple monomial
AT	idempotent and commutative	FO^1	$\mathrm{AT} \subsetneq \mathrm{RExt}_A(\mathrm{AT}) \subsetneq \mathrm{RExt}_A(\mathrm{PT})$ L is AT-measurable iff L or $\overline{L}$ contains $\bigcap_{a \in A} A^* a A^*$