Test sets of commutative languages

Štěpán Holub

Department of Algebra
Charles University in Prague

JM'06, 29th August, Rennes

Commutative languages

Commutative closure of a word $u = \ell_1 \dots \ell_n$ is the language

$$c(u) = \{\ell_{\sigma(1)} \dots \ell_{\sigma(n)} \mid \sigma \in S_n\}.$$

Commutative closure of a language is

$$c(L) = \bigcup_{u \in L} c(u).$$

Language L is said to be commutative iff

$$L = c(L).$$

Test sets

Two morphisms agree on a language L iff

$$g(u) = h(u)$$

for any $u \in L$. Write

$$g \equiv_L h$$

A subset T of the language L is called its test set iff for any two morphisms g, h

$$g \equiv_L h \quad \Leftrightarrow \quad g \equiv_T h.$$

Parikh vectors

Parikh vector of a word $u = a_1^{k_1} \dots a_n^{k_n}$ is the n-tuple $(k_1, \dots, k_n)$.

- A commutative language is given by the set of its Parikh vectors
- A basis of the vector space over $\mathbb{Q}$ spanned by the Parikh vectors of a language is called Parikh basis of the language
- If two morphisms agree lengthwise on a Parikh basis, they agree lengthwise on the whole language.

Parikh vectors

Parikh vector of a word $u = a_1^{k_1} \dots a_n^{k_n}$ is the n-tuple $(k_1, \dots, k_n)$.

- A commutative language is given by the set of its Parikh vectors
- A basis of the vector space over $\mathbb{Q}$ spanned by the Parikh vectors of a language is called Parikh basis of the language
- If two morphisms agree lengthwise on a Parikh basis, they agree lengthwise on the whole language.

Known bounds

Theorem (Hakala, Kortelainen 1997)

Any commutative language over n letters has a test of cardinality at most $3n^2$.

There is a commutative language over n letters, the smallest test set of which has cardinality at least $\frac{1}{9}n^2$.

Theorem (Holub, Kortelainen 2001)

The commutative language $c(a_1^{k_1} \cdots a_n^{k_n})$ has a test of cardinality at most $10n$.

Each test set of the commutative language $c(a_1 \cdots a_n)$ has cardinality at least $n - 1$.

Known bounds

Theorem (Hakala, Kortelainen 1997)

Any commutative language over n letters has a test of cardinality at most $3n^2$.

There is a commutative language over n letters, the smallest test set of which has cardinality at least $\frac{1}{9}n^2$.

Theorem (Holub, Kortelainen 2001)

The commutative language $c(a_1^{k_1} \cdots a_n^{k_n})$ has a test of cardinality at most $10n$.

Each test set of the commutative language $c(a_1 \cdots a_n)$ has cardinality at least $n - 1$.

Known bounds

Theorem (Hakala, Kortelainen 1997)

Any commutative language over n letters has a test of cardinality at most $3n^2$.

There is a commutative language over n letters, the smallest test set of which has cardinality at least $\frac{1}{9}n^2$.

Theorem (Holub, Kortelainen 2001)

The commutative language $c(a_1^{k_1} \cdots a_n^{k_n})$ has a test of cardinality at most $10n$.

Each test set of the commutative language $c(a_1 \cdots a_n)$ has cardinality at least $n-1$.

Known bounds

Theorem (Hakala, Kortelainen 1997)

Any commutative language over n letters has a test of cardinality at most $3n^2$.

There is a commutative language over n letters, the smallest test set of which has cardinality at least $\frac{1}{9}n^2$.

Theorem (Holub, Kortelainen 2001)

The commutative language $c(a_1^{k_1} \cdots a_n^{k_n})$ has a test of cardinality at most $10n$.

Each test set of the commutative language $c(a_1 \cdots a_n)$ has cardinality at least $n - 1$.

The Main Question

Commutative language can be

- simple: it has a Parikh test set of size $\mathcal{O}(n)$
- complicated: it has only Parikh test sets of size $\Omega(n^2)$

What is the criterion?

The Main Question

Commutative language can be

- simple: it has a Parikh test set of size $\mathcal{O}(n)$
- complicated: it has only Parikh test sets of size $\Omega(n^2)$

What is the criterion?

Parikh test set

A set $T \subset L$ is a Parikh test set of L iff $c(T)$ is a test set of $c(L)$.

Two sources of the size of a test sets:

- Large Parikh test sets
- Many words with the same Parikh vector

Parikh test set

A set $T \subset L$ is a Parikh test set of L iff $c(T)$ is a test set of $c(L)$.

Two sources of the size of a test sets:

- Large Parikh test sets
- Many words with the same Parikh vector

Parikh test set

A set $T \subset L$ is a Parikh test set of L iff $c(T)$ is a test set of $c(L)$.

Two sources of the size of a test sets:

- Large Parikh test sets
- Many words with the same Parikh vector

Parikh test sets and classical test sets

A set $T \subset L$ is a Parikh test set of L iff $c(T)$ is a test set of $c(L)$.

- Parikh test sets give a lower bound to cardinality of test sets
- If T is a Parikh test set of L, then L has a test set of cardinality $10 \parallel T \parallel$, where

$$\parallel T \parallel = \sum_{t \in T} |t|.$$

A complicated language

- $X = \{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n\}$, $|X| = 3n$.
- $L = c(\{a_i b_j c_i \mid i, j = 1, \ldots, n\})$, $|L| = (6)n^2$

$$T \subset L, \qquad |T| < n^2 \Rightarrow c(a_n b_n c_n) \cap T = \emptyset$$

$$\begin{array}{ccc}
g(a_n) = a^2 & g(b_n) = b & h(c_n) = a^2 \\
g(a_n) = a & g(b_n) = b & g(c_n) = a \\
h(a_j) = a & h(b_j) = a & g(c_j) = a \\
h(a_j) = a & h(b_j) = a & h(c_j) = a
\end{array}$$

A complicated language

- $X = \{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n\}$, $|X| = 3n$.
- $L = c(\{a_i b_j c_i \mid i, j = 1, \ldots, n\})$, $|L| = (6)n^2$

$$T \subset L, \qquad |T| < n^2 \Rightarrow c(a_n b_n c_n) \cap T = \emptyset$$

$$\begin{array}{ccc} g(a_n) = a^2 & g(b_n) = b & h(c_n) = a^2 \\ g(a_n) = a & g(b_n) = b & g(c_n) = a \\ h(a_j) = a & h(b_j) = a & g(c_j) = a \\ h(a_j) = a & h(b_j) = a & h(c_j) = a \end{array}$$

A complicated language

- $X = \{a_1, \ldots, a_n, b_1, \ldots, b_n, c_1, \ldots, c_n\}$, $|X| = 3n$.
- $L = c(\{a_i b_j c_i \mid i, j = 1, \ldots, n\})$, $|L| = (6)n^2$

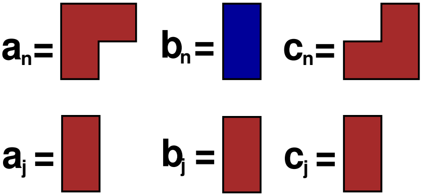
$$T \subset L, \qquad |T| < n^2 \Rightarrow c(a_n b_n c_n) \cap T = \emptyset$$

$$\begin{array}{ccc}
g(a_n) = a^2 & g(b_n) = b & h(c_n) = a^2 \\
g(a_n) = a & g(b_n) = b & g(c_n) = a \\
h(a_j) = a & h(b_j) = a & g(c_j) = a \\
h(a_j) = a & h(b_j) = a & h(c_j) = a
\end{array}$$

A complicated language

$$L = c(\{a_i b_j c_i \mid i,j = 1,\ldots,n\}); \qquad c(a_n b_n c_n) \cap T = \emptyset$$

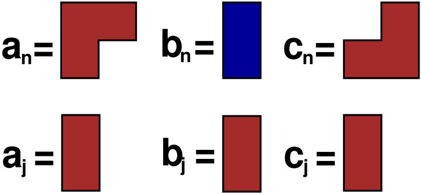
$$\begin{array}{ccc} g(a_n) = a^2 & g(b_n) = b & g(c_n) = a \\ h(a_n) = a & h(b_n) = b & h(c_n) = a^2 \\ g(a_j) = a & g(b_j) = a & g(c_j) = a \\ h(a_j) = a & h(b_j) = a & h(c_j) = a \end{array}$$



Many Parikh vectors??

A complicated language

$$L = c(\{a_i b_j c_i \mid i,j = 1,\ldots,n\}); \qquad c(a_n b_n c_n) \cap T = \emptyset$$



Many Parikh vectors??

A simple language with many Parikh vectors

- $X_1 = \{a_1, \dots, a_n, b_1, \dots, b_n\}$, $|X| = 2n$.
- $L_1 = c(\{a_i b_j \mid i, j = 1, \dots, n\})$, $|L| = 2n^2$

$$T = \{a_n b_i \mid i = 1, \dots, n\} \bigcup \{a_i b_n \mid i = 1, \dots, n\}$$

is a test set of L with $|T| = 2n - 1$, and the Parikh size n.

A simple language with many Parikh vectors

- $X_1 = \{a_1, \ldots, a_n, b_1, \ldots, b_n\}$, $|X| = 2n$.
- $L_1 = c(\{a_i b_j \mid i, j = 1, \ldots, n\})$, $|L| = 2n^2$

$$T = \{a_n b_i \mid i = 1, \ldots, n\} \bigcup \{a_i b_n \mid i = 1, \ldots, n\}$$

is a test set of L with $|T| = 2n - 1$, and the Parikh size n.

Coincidence graph

Coincidence graph of L:

- G(L)=(X,E)
  - X is alphabet
  - $(a, b) \in E$ iff a and b occur together in a word.
- Undirected
- Loops are allowed

Difference support

Example: Let $a_i \in L$ for each letter a_i. Then the coincidence graph is irrelevant.

We say that $D \subset X$ is a difference support of L if there exist morphisms g and h, which are lenght equivavelnt on L, and

$$|g(a)| \neq |h(a)| \qquad \text{iff} \qquad a \in D.$$

Difference support

Example: Let $a_i \in L$ for each letter a_i. Then the coincidence graph is irrelevant.

We say that $D \subset X$ is a difference support of L if there exist morphisms g and h, which are lenght equivavelnt on L, and

$$|g(a)| \neq |h(a)| \qquad \text{iff} \qquad a \in D.$$

Difference support

Example: Let $a_i \in L$ for each letter a_i. Then the coincidence graph is irrelevant.

We say that $D \subset X$ is a difference support of L if there exist morphisms g and h, which are lenght equivavelnt on L, and

$$|g(a)| \neq |h(a)| \qquad \text{iff} \qquad a \in D.$$

Difference support

Example: Let $a_i \in L$ for each letter a_i. Then the coincidence graph is irrelevant.

We say that $D \subset X$ is a difference support of L if there exist morphisms g and h, which are lenght equivavelnt on L, and

$$|g(a)| \neq |h(a)| \qquad \text{iff} \qquad a \in D.$$

In the example the only difference support is the empty set.

Difference support

Example: Let $a_i \in L$ for each letter a_i. Then the coincidence graph is irrelevant.

We say that $D \subset X$ is a difference support of L if there exist morphisms g and h, which are lenght equivavelnt on L, and

$$|g(a)| \neq |h(a)| \qquad \text{iff} \qquad a \in D.$$

The difference support is given by the basis of Parikh vectors.

$$\begin{aligned} B \cdot d &= 0 \\ d &= (d_1, \ldots, d_n) \\ D &= \{a_i \mid d_i \neq 0\} \end{aligned}$$

The main claim

Theorem: If $G(L)$, $G(T)$, and Parikh basis of L are effectively given, then it is decidable whether T is a Parikh test set of L.

Note: Difference supports can be effectivelly found.

Parikh test set criterion

We say that a and b are D-connected, $D \subset X$,
iff there exists a path from a to b, which do not contain two consecutive vertices from $X \setminus D$.

Theorem:
$T \subset L$ is a Parikh test set iff T contains a Parikh basis of L for each difference support D the graph $G(T)$ satisfies:

- a and b are D-connected in $G(L) \Rightarrow a$ and b are D-connected in $G(T)$
- For a letter $a \in D$ there is a path from a to a of odd length in $G(L) \Rightarrow$ there is such a path in $G(T)$

Parikh test set criterion

We say that a and b are D-connected, $D \subset X$,
iff there exists a path from a to b, which do not contain two consecutive vertices from $X \setminus D$.

Theorem:
$T \subset L$ is a Parikh test set iff T contains a Parikh basis of L for each difference support D the graph $G(T)$ satisfies:

- a and b are D-connected in $G(L) \Rightarrow a$ and b are D-connected in $G(T)$
- For a letter $a \in D$ there is a path from a to a of odd length in $G(L) \Rightarrow$ there is such a path in $G(T)$

Parikh test set criterion

We say that a and b are D-connected, $D \subset X$, iff there exists a path from a to b, which do not contain two consecutive vertices from $X \setminus D$.

Theorem:
$T \subset L$ is a Parikh test set iff T contains a Parikh basis of L for each difference support D the graph $G(T)$ satisfies:

- a and b are D-connected in $G(L) \Rightarrow a$ and b are D-connected in $G(T)$
- For a letter $a \in D$ there is a path from a to a of odd length in $G(L) \Rightarrow$ there is such a path in $G(T)$

Parikh test set criterion

We say that a and b are D-connected, $D \subset X$, iff there exists a path from a to b, which do not contain two consecutive vertices from $X \setminus D$.

Theorem:
$T \subset L$ is a Parikh test set iff T contains a Parikh basis of L for each difference support D the graph $G(T)$ satisfies:

- a and b are D-connected in $G(L)$ $\Rightarrow$ a and b are D-connected in $G(T)$
- For a letter $a \in D$ there is a path from a to a of odd length in $G(L)$ $\Rightarrow$ there is such a path in $G(T)$

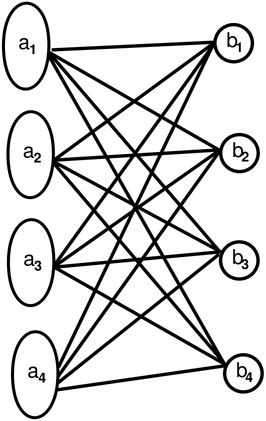
Parikh test set criterion

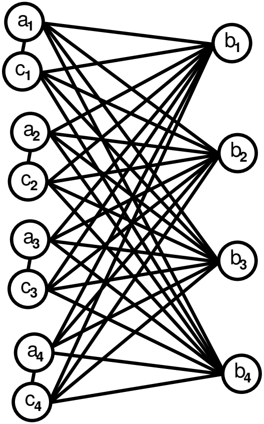
We say that a and b are D-connected, $D \subset X$, iff there exists a path from a to b, which do not contain two consecutive vertices from $X \setminus D$.

Theorem:
$T \subset L$ is a Parikh test set iff T contains a Parikh basis of L for each difference support D the graph $G(T)$ satisfies:

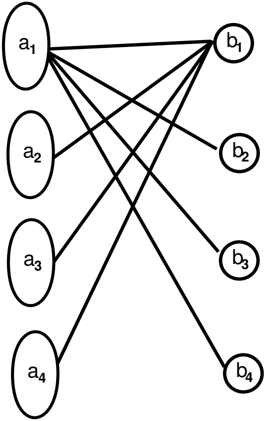
- a and b are D-connected in $G(L) \Rightarrow a$ and b are D-connected in $G(T)$
- For a letter $a \in D$ there is a path from a to a of odd length in $G(L) \Rightarrow$ there is such a path in $G(T)$

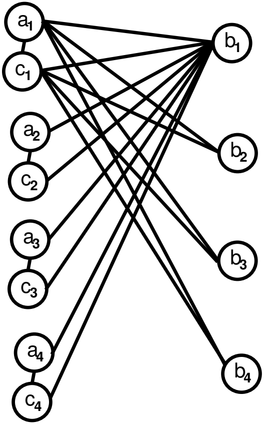
Examples



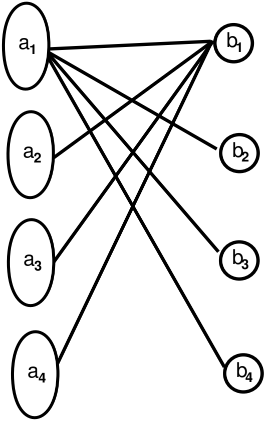


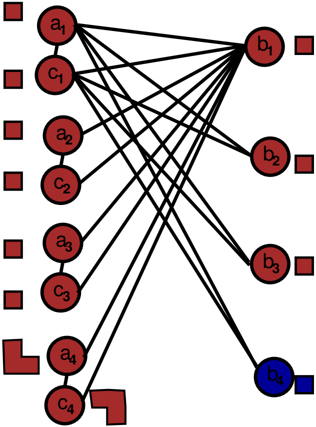
Examples



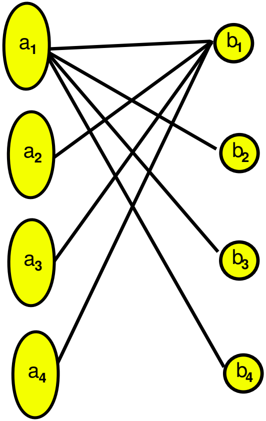


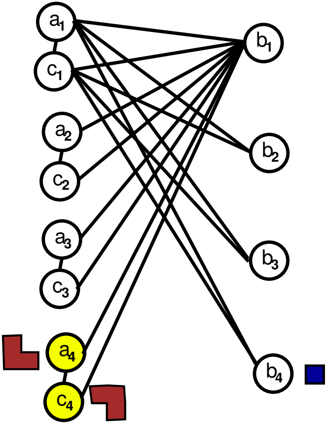
Examples





Examples





Why cycles of odd length are important

X={a,b,c,d}

$$g(a) = rsr \quad g(b) = s \quad g(c) = rsr \quad g(d) = s$$
$$h(a) = r \quad h(b) = srs \quad h(c) = r \quad h(d) = srs$$

g and h agree on

$$c\{ab, bc, cd, da\}$$

but not on

$$c\{ab, bc, ca\}$$