Monoids and Maximal Codes

Fabio Burderi

Dipartimento di Matematica e Informatica
Università degli studi di Palermo,
burderi@math.unipa.it

WORDS, Sept. 12-16 2011, Prague

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Attention !!

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^+ *messages*.

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Attention !!

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^+ *messages*.

- If $w \in A^*$, a *factorization* of w is a sequence of words $(v_i)_{1 \leq i \leq s}$ such that $w = v_1 v_2 \cdots v_s$.

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Attention !!

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^+ *messages*.

- If $w \in A^*$, a *factorization* of w is a sequence of words $(v_i)_{1 \leq i \leq s}$ such that $w = v_1 v_2 \cdots v_s$.
- If X is a code, a *relation* between code words is a pair of factorizations $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ into code words of a same message $w \in X^+$; the relation is said non-trivial if the factorizations are distinct.

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Attention !!

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^+ *messages*.

- If $w \in A^*$, a *factorization* of w is a sequence of words $(v_i)_{1 \leq i \leq s}$ such that $w = v_1 v_2 \cdots v_s$.
- If X is a code, a *relation* between code words is a pair of factorizations $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ into code words of a same message $w \in X^+$; the relation is said non-trivial if the factorizations are distinct.
- We say that the relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ is *prime* if for all $i < s$ and for all $j < t$ one has $x_1 x_2 \cdots x_i \neq y_1 y_2 \cdots y_j$.

Let A be an alphabet. Let A^* denote the free monoid generated by A, and let $A^+ = A^* \setminus \{\varepsilon\}$.

Attention !!

A *code* X over A is a subset of A^+. The words of X are called *code words*, the elements of X^+ *messages*.

- If $w \in A^*$, a *factorization* of w is a sequence of words $(v_i)_{1 \leq i \leq s}$ such that $w = v_1 v_2 \cdots v_s$.
- If X is a code, a *relation* between code words is a pair of factorizations $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ into code words of a same message $w \in X^+$; the relation is said non-trivial if the factorizations are distinct.
- We say that the relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$ is *prime* if for all $i < s$ and for all $j < t$ one has $x_1 x_2 \cdots x_i \neq y_1 y_2 \cdots y_j$.
- A relation $w = x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$, can be *univocally* factorized into prime relations.

A code X is said to be *uniquely decipherable* (*UD*) if there are not non-trivial relations on X.
Every message has an unique factorization into code words:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m, \qquad x_i, y_j \in X$$

implies $\quad n = m \quad$ and $\quad x_1 = y_1, \ldots, x_n = y_n.$

A code X is said to be *uniquely decipherable* (*UD*) if there are not non-trivial relations on X.
Every message has an unique factorization into code words:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m, \qquad x_i, y_j \in X$$

implies $\quad n = m \quad$ and $\quad x_1 = y_1, \ldots, x_n = y_n.$

Example 1 $\quad A = \{0, 1\}, \quad X = A^2 = \{00, 01, 10, 11\},$

$$z = 0100101011$$

A code X is said to be *uniquely decipherable* (*UD*) if there are not non-trivial relations on X.
Every message has an unique factorization into code words:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m, \qquad x_i, y_j \in X$$

implies $\quad n = m \quad$ and $\quad x_1 = y_1, \ldots, x_n = y_n.$

Example 1 $\quad A = \{0, 1\}, \quad X = A^2 = \{00, 01, 10, 11\},$

$$z = 0100101011 = 01 \cdot 00 \cdot 10 \cdot 10 \cdot 11$$

A code X is said to be *uniquely decipherable* (*UD*) if there are not non-trivial relations on X.
Every message has an unique factorization into code words:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m, \qquad x_i, y_j \in X$$

implies $n = m$ and $x_1 = y_1, \ldots, x_n = y_n$.

Example 1 $A = \{0, 1\}, \quad X = A^2 = \{00, 01, 10, 11\},$

$$z = 0100101011 = 01 \cdot 00 \cdot 10 \cdot 10 \cdot 11$$

Example 2 $X = \{0, 01, 10\}$

$$z = 010$$

A code X is said to be *uniquely decipherable* (*UD*) if there are not non-trivial relations on X.
Every message has an unique factorization into code words:

$$x_1 x_2 \cdots x_n = y_1 y_2 \cdots y_m, \qquad x_i, y_j \in X$$

implies $n = m$ and $x_1 = y_1, \ldots, x_n = y_n$.

Example 1 $A = \{0, 1\}, \quad X = A^2 = \{00, 01, 10, 11\},$

$$z = 0100101011 = 01 \cdot 00 \cdot 10 \cdot 10 \cdot 11$$

Example 2 $X = \{0, 01, 10\}$

$$z = 010 = 0 \cdot 10 = 01 \cdot 0$$

Let X be a code and let $P = \{X_i \mid i \in I\}$ be a partition of X i.e. :
$\bigcup_{i \in I} X_i = X$ and $X_i \cap X_j = \emptyset$, iff $i \neq j$.

Let X be a code and let $P = \{X_i \mid i \in I\}$ be a partition of X i.e. :
$\bigcup_{i \in I} X_i = X$ and $X_i \cap X_j = \emptyset$, iff $i \neq j$.

A *P-factorization* of a message $w \in X^+$ is a factorization

$$w = z_1 z_2 \cdots z_t$$

where:

- for each i $z_i \in X_k^+$, for some $X_k \in P$
- if $t > 1$, $z_i \in X_k^+ \Rightarrow z_{i+1} \notin X_k^+$, $(1 \leq i \leq t-1)$.

Example 3 $X = \{00, 11, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}, \quad X_1 = \{00, 11\}, \quad X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+, \quad w = 11 \cdot 00 \cdot 000 \cdot 111$

Example 3 $X = \{00, 11, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}, \quad X_1 = \{00, 11\}, \quad X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+, \quad w = 11 \cdot 00 \cdot 000 \cdot 111$

$w = z_1 z_2 = (11 \cdot 00)(000 \cdot 111)$

Example 3 $X = \{00, 11, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}, \quad X_1 = \{00, 11\}, \quad X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+, \quad w = 11 \cdot 00 \cdot 000 \cdot 111$

$w = z_1 z_2 = (11 \cdot 00)(000 \cdot 111)$

$w = 11 \cdot 000 \cdot 00 \cdot 111 \quad w = u_1 u_2 u_3 u_4 = (11)(000)(00)(111)$

$z_1 z_2$ and $u_1 u_2 u_3 u_4$ are $P -$ *factorizzations* of z.

Example 3 $X = \{00, 11, 000, 111\} = \{x_1, x_2, x_3, x_4\}$,

$P = \{X_1, X_2, \}, \quad X_1 = \{00, 11\}, \quad X_2 = \{000, 111\}$. Let

$w = 1100000111 \in X^+, \quad w = 11 \cdot 00 \cdot 000 \cdot 111$

$w = z_1 z_2 = (11 \cdot 00)(000 \cdot 111)$

$w = 11 \cdot 000 \cdot 00 \cdot 111 \quad w = u_1 u_2 u_3 u_4 = (11)(000)(00)(111)$

$z_1 z_2$ and $u_1 u_2 u_3 u_4$ are $P -$ *factorizzations* of z.

The partition P is called a *coding partition* if any element $w \in X^+$ has a *unique P-factorization*, i.e. if

$$w = z_1 z_2 \cdots z_s = u_1 u_2 \cdots u_t,$$

with $z_1 z_2 \cdots z_s, \quad u_1 u_2 \cdots u_t$ *P-factorizations* of w, then:
$s = t$ and $z_i = u_i$ for $i = 1, \ldots, s$.

Example 3 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$, $X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$. $w = 1100000111 = 11 \cdot 00000 \cdot 111$

P is a coding partition of X.

Example 3 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$, $X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$. $w = 1100000111 = 11 \cdot 00000 \cdot 111$

P is a coding partition of X.

- Let $P = \{X_i \mid i \in I\}$ be a partition of a code X. The partition P is a coding partition iff for every prime relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$, the code words x_i, y_j belong to the same component of the partition.

Example 3 $X = \{00, 11, 000, 111\}$, $P = \{X_1, X_2, \}$, $X_1 = \{00, 000\}$, $X_2 = \{11, 111\}$. $w = 1100000111 = 11 \cdot 00000 \cdot 111$

P is a coding partition of X.

- Let $P = \{X_i \mid i \in I\}$ be a partition of a code X. The partition P is a coding partition iff for every prime relation $x_1 x_2 \cdots x_s = y_1 y_2 \cdots y_t$, the code words x_i, y_j belong to the same component of the partition.
- A code X is called *ambiguous* if it is not *UD*.
- A code is called *totally ambiguous* (*TA*) if $|X| > 1$ and the only coding partition is the trivial partition: $P = \{X\}$.

Example 2 $X = \{0, 01, 10\}$.

The word $w = 010 \in X^+$ has two factorizations : $w = 0 \cdot 10 = 01 \cdot 0$.

X is a totally ambiguous code.

Given a code $X \subseteq A^*$ we can study the properties of the monoid $M = X^*$.

Given a code $X \subseteq A^*$ we can study the properties of the monoid $M = X^*$.

Let M be a monoid generated by submonoids $M_\lambda, \lambda \in \Lambda$, and let $m \in M$. An expression of m of the form $m_1 m_2 \cdots m_r$, where $r \geq 0$, $1 \neq m_i \in M_{\lambda_i}$, $\lambda_i \neq \lambda_{i+1}$, is said in *reduced form* with respect to M_λ's.

Given a code $X \subseteq A^*$ we can study the properties of the monoid $M = X^*$.

Let M be a monoid generated by submonoids $M_\lambda, \lambda \in \Lambda$, and let $m \in M$. An expression of m of the form $m_1 m_2 \cdots m_r$, where $r \geq 0$, $1 \neq m_i \in M_{\lambda_i}$, $\lambda_i \neq \lambda_{i+1}$, is said in *reduced form* with respect to M_λ's.

- M is the free product of the M_λ's iff every element of M has an unique expression in reduced form with respect to M_λ's

Given a code $X \subseteq A^*$ we can study the properties of the monoid $M = X^*$.

Let M be a monoid generated by submonoids $M_\lambda, \lambda \in \Lambda$, and let $m \in M$. An expression of m of the form $m_1 m_2 \cdots m_r$, where $r \geq 0$, $1 \neq m_i \in M_{\lambda_i}$, $\lambda_i \neq \lambda_{i+1}$, is said in *reduced form* with respect to M_λ's.

- M is the free product of the M_λ's iff every element of M has an unique expression in reduced form with respect to M_λ's

- A family $\{M_\lambda \mid \lambda \in \Lambda\}$ of submonoids of M is a *free factorization* of M if M is the free product of the M_λ's. The M_λ's are called the *free factors* of the free factorization; moreover we say that a monoid M is *freely indecomposable* if M cannot be expressed as a free product of nontrivial monoids.

Remark: a free factor is not, in general, a free monoid.

- A *UD* code $X \subseteq A^+$ is said to be a *maximal UD* code if X is not properly contained in any other *UD* code over A.

  For example uniform codes A^n are maximal *UD* codes $\forall n \geq 1$.

- A *UD* code $X \subseteq A^+$ is said to be a *maximal UD* code if X is not properly contained in any other *UD* code over A.

  For example uniform codes A^n are maximal *UD* codes $\forall n \geq 1$.

- Any *UD* code $X \subseteq A^+$ is contained in some maximal *UD* code over A.

We introduce now a binary relation $\preceq$ on the set of submonoids of A^*.

We introduce now a binary relation $\preceq$ on the set of submonoids of A^*.

- Let $M, N \subseteq A^*$ be monoids we say that $N \preceq M$ if there exists a monoid $L \subseteq A^*$ such that $M = N * L$.

We introduce now a binary relation $\preceq$ on the set of submonoids of A^*.

- Let $M, N \subseteq A^*$ be monoids we say that $N \preceq M$ if there exists a monoid $L \subseteq A^*$ such that $M = N * L$.
- The relation $\preceq$ is a partial order on the set of submonoids of A^*.

We introduce now a binary relation $\preceq$ on the set of submonoids of A^*.

- Let $M, N \subseteq A^*$ be monoids we say that $N \preceq M$ if there exists a monoid $L \subseteq A^*$ such that $M = N * L$.
- The relation $\preceq$ is a partial order on the set of submonoids of A^*.

Theorem

Any submonoid $N \subseteq A^*$ is contained in a submonoid $M \subseteq A^*$ such that: $N \preceq M$ and M is maximal with respect to the partial order $\preceq$.

Definition

We say that a submonoid M of A^* is *full* if it is maximal with respect to the partial order $\preceq$.

Definition

We say that a submonoid M of A^* is *full* if it is maximal with respect to the partial order $\preceq$.

Definition

A code $X \subseteq A^+$ is said *maximal* if the monoid X^* is full.

Definition

We say that a submonoid M of A^* is *full* if it is maximal with respect to the partial order $\preceq$.

Definition

A code $X \subseteq A^+$ is said *maximal* if the monoid X^* is full.

- Let $M \subseteq A^*$ be a monoid. If M is maximal with respect to the inclusion order $\subseteq$ then it is full.

Definition

We say that a submonoid M of A^* is *full* if it is maximal with respect to the partial order $\preceq$.

Definition

A code $X \subseteq A^+$ is said *maximal* if the monoid X^* is full.

- Let $M \subseteq A^*$ be a monoid. If M is maximal with respect to the inclusion order $\subseteq$ then it is full.

A free monoid $M \subseteq A^*$ is said *maximal free* if $M \neq A^*$ and M is not properly contained in any other free monoid different from A^*.

Definition

We say that a submonoid M of A^* is *full* if it is maximal with respect to the partial order $\preceq$.

Definition

A code $X \subseteq A^+$ is said *maximal* if the monoid X^* is full.

- Let $M \subseteq A^*$ be a monoid. If M is maximal with respect to the inclusion order $\subseteq$ then it is full.

A free monoid $M \subseteq A^*$ is said *maximal free* if $M \neq A^*$ and M is not properly contained in any other free monoid different from A^*.

- Let M be a free monoid. If M is maximal free then it is full.

- If X is a UD code then the monoid X^* is free.

- If X is a UD code then the monoid X^* is free.

A code X is a base if X is the minimal set of generators of X^* i.e. if no word on X is a concatenation of other code words.

- If X is a *UD* code then the monoid X^* is free.

A code X is a base if X is the minimal set of generators of X^* i.e. if no word on X is a concatenation of other code words.

Theorem

Let $X \subseteq A^+$ be a code that is a base. Then X is a maximal UD code iff X^* is a full and free submonoid of A^*.

- A word $w \in A^*$ is a *factor* of a word $z \in A^*$ if there exist $u, v \in A^*$ such that $z = uwv$.
  For any $X \subseteq A^*$ let $F(X)$ denote the set of factors of words in X.

- A word $w \in A^*$ is a *factor* of a word $z \in A^*$ if there exist $u, v \in A^*$ such that $z = uwv$.
  For any $X \subseteq A^*$ let $F(X)$ denote the set of factors of words in X.
- A set $X \subseteq A^*$ is called *dense* if $F(X) = A^*$. A set that is not dense is called *thin*.

- A word $w \in A^*$ is a *factor* of a word $z \in A^*$ if there exist $u, v \in A^*$ such that $z = uwv$.
  For any $X \subseteq A^*$ let $F(X)$ denote the set of factors of words in X.
- A set $X \subseteq A^*$ is called *dense* if $F(X) = A^*$. A set that is not dense is called *thin*.
- A set $X \subseteq A^*$ is called *complete* if X^* is dense.

Theorem

Any full monoid $M \subseteq A^*$ is dense in A^*.

Theorem

Any full monoid $M \subseteq A^*$ is dense in A^*.

Corollary

If $X \subseteq A^+$ is a maximal code then it is a complete set.

Theorem

Any full monoid $M \subseteq A^*$ is dense in A^*.

Corollary

If $X \subseteq A^+$ is a maximal code then it is a complete set.

- The inverse of previous theorem is not true.
  Let $A = \{a, b\}$ and let M be the submonoid of A^* composed of the words on A^* having as many a's as b's. The base of M is a maximal UD code, it is denoted by D and it is called the Dyck code over A.

Theorem

Any full monoid $M \subseteq A^*$ is dense in A^*.

Corollary

If $X \subseteq A^+$ is a maximal code then it is a complete set.

- The inverse of previous theorem is not true.
  Let $A = \{a, b\}$ and let M be the submonoid of A^* composed of the words on A^* having as many a's as b's. The base of M is a maximal UD code, it is denoted by D and it is called the Dyck code over A. Indeed D is dense and for each $x \in D$ the code $D \setminus \{x\}$ remains dense but it is no more a maximal UD code and so $(D \setminus \{x\})^*$ it is not full in A^*.

Lemma (Schützenberger)

Let $X \subseteq A^+$ be a regular and complete code. Then there exist a word $v \in X^+$ and a positive integer m such that for any word $w \in A^*$, $(vwv)^m \in X^+$.

Lemma (Schützenberger)

Let $X \subseteq A^+$ be a regular and complete code. Then there exist a word $v \in X^+$ and a positive integer m such that for any word $w \in A^*$, $(vwv)^m \in X^+$.

Theorem

Let X be a regular code. Then X is complete iff X is a maximal code.

Lemma (Schützenberger)

Let $X \subseteq A^+$ be a regular and complete code. Then there exist a word $v \in X^+$ and a positive integer m such that for any word $w \in A^*$, $(vwv)^m \in X^+$.

Theorem

Let X be a regular code. Then X is complete iff X is a maximal code.

Theorem

Every regular code is contained in a maximal regular code.

Thank you for your attention!