Recognizers

Automata, Minimal recognizer
Recognizable set
Syntactic monoid
Rational decomposition of recognizable sets
Prefix decomposition of recognizable set
The pumping lemma and size of recognizable set
Automata or recognizer

\[ M = (Q, \Sigma, F) \]

\[ i \in Q \text{ Initial state} \]

\[ T \subseteq Q \text{ Terminal states} \]
Recognizing word. Behavior

\[ \alpha \in \Sigma^* \quad iF_\alpha \in Q \text{ or } iF_\alpha = \emptyset \]

\[ iF_\alpha \in T. \]

\[ \mathcal{M} \text{ recognizes } \alpha \]

\[ \overline{\mathcal{M}} = \{ \alpha \in \Sigma^* \mid iF_\alpha \in T \} \text{ behaviour} \]
Notation

\[ A \subseteq \Sigma^*, \ B \subseteq \Sigma^* \quad A \neq \emptyset, \ B \neq \emptyset \]

\[ A \cdot B = \{ \alpha \in \Sigma^* | \alpha = ab; \ a \in A, \ b \in B \}, \]
\[ A^+ = \{ \alpha \in \Sigma^* | \alpha = a_1 \cdot \ldots \cdot a_n; \ a_i \in A, \ 1 \leq i \leq n, \ n > 0 \}, \]
\[ A^* = A^+ \cup \{ \Lambda \}. \]
Examples

| M | = \Sigma^*.

{0}^+
Completion. Theorem

\[ \mathcal{M} = (Q, \Sigma, F) \]  
\[ i \in Q \text{ and } T \subseteq Q \]  
\[ \text{completion } \mathcal{M}^c \text{ of } \mathcal{M} \]  
\[ \mathcal{M}^c = (\mathcal{M}^c, i, T) \]  
\[ |\mathcal{M}^c| = |\mathcal{M}|. \]
Recognizable set

subset $A$ of $\Sigma^*$

exists a recognizer $M$

$A = |M|$

recognizable
\( \mathcal{M} = (M, i, T) \) where \( M = (Q, \Sigma, F) \)

\( R = \{ iF_\alpha \mid \alpha \in \Sigma^* \} \)

\( R = Q \)

\( M^a = (R, \Sigma, F^a) \)

Accessible part

\( M^a = (M^a, i, T) \)

\( |M^a| = |M| \)
Co accessible

$\mathcal{M} = (M, i, T)$ where $M = (Q, \Sigma, F)$

$S = \{q \mid qF_\alpha \in T \text{ for some } \alpha \in \Sigma^*\}$.

$S = Q$ we call $\mathcal{M}$ coaccessible.

both accessible and coaccessible.

$M^b = (S, \Sigma, F^b)$, coaccessible part

$M^b = (M^b, i, T)$

$qF^b_\alpha = qF_\alpha$ for $q \in S$, $\alpha \in \Sigma^*$

$|M^b| = |M|$
Example
\[ M = (Q, \Sigma, F) \text{ complete} \]

\[ q * \alpha = q F_\alpha \]

\[ A \subseteq \Sigma^*, \ S \subseteq Q \]

\[ q * A = \{ q * \alpha \mid \alpha \in A \} \]

\[ S * \alpha = \{ q * \alpha \mid q \in S \} \]

\[ S * A = \{ q * \alpha \mid q \in S, \ \alpha \in A \}. \]

\[ q * \alpha^{-1} = \{ p \in Q \mid q = p * \alpha \} \]

\[ q * A^{-1} = \{ p \in Q \mid q = p * \alpha \text{ for some } \alpha \in A \} \]

\[ S * \alpha^{-1} = \{ p \in Q \mid p * \alpha \in S \} \]

\[ S * A^{-1} = \{ p \in Q \mid p * \alpha \in S \text{ for some } \alpha \in A \}. \]

\[ q^{-1} \circ p = \{ \alpha \in \Sigma^* \mid p = q * \alpha \} \]

\[ q^{-1} \circ R = \{ \alpha \in \Sigma^* \mid q * \alpha \in R \} \]

\[ S^{-1} \circ R = \{ \alpha \in \Sigma^* \mid q * \alpha \in R \text{ for some } q \in S \}. \]
Propositions

Let $\mathcal{M} = (Q, \Sigma, F)$ be a state machine, $A, B, C \subseteq \Sigma^*$ and $S \subseteq Q$.

(i) $(S \cdot A) \cdot B = S \cdot (A \cdot B)$

(ii) $(S \cdot A^{-1}) \cdot B^{-1} = S \cdot (B \cdot A)^{-1}$

Let $\mathcal{M} = (Q, \Sigma, F)$ and $\mathcal{M} = (\mathcal{M}, i, T)$ and $A = |\mathcal{M}|$ then

$A = i^{-1} \circ T$.

If $q = i \cdot \alpha$, $\alpha \in \Sigma^*$, then

$q^{-1} \circ T = \alpha^{-1} A$. 
Proposition

Let $M = (Q, \Sigma, F)$ and $N = (M, i, T)$, then $N$

is accessible if and only if $Q = i \ast (\Sigma^*)$

and $N$ is coaccessible if and only if $Q = T \ast (\Sigma^*)^{-1}$. 
Minimal recognizer

\[ A \subseteq \Sigma^* \quad A \text{ is recognizable} \]

\[ \mathcal{M} = (M, i, T) \quad \text{where } M = (Q, \Sigma, F) \quad \text{and} \quad A = |\mathcal{M}| \]

\[ \alpha^{-1} \cdot A = \{ \beta \in \Sigma^* \mid \alpha \beta \in A \} \Rightarrow Q_A = \{ \alpha^{-1} \cdot A \mid \alpha \in \Sigma^* \} \]

\[ F^A : Q_A \times \Sigma \rightarrow Q_A \quad (\alpha^{-1} \cdot A) F^A_\sigma = (\alpha \sigma)^{-1} \cdot A \]

\[ \emptyset F^A = \emptyset \]

\[ M_A = (Q_A, \Sigma, F^A) \]

\[ i_A = A \]

\[ T_A = \{ \alpha^{-1} \cdot A \in Q_A \mid \alpha \in A \} \]

\[ \mathcal{M}_A = (M_A, i_A, T_A) \]

If \( A \subseteq \Sigma^* \) is recognizable then \( \mathcal{M}_A \) is a recognizer with the property that \( |\mathcal{M}_A| = A \).
\[ A = \{0\} \cdot \{1\}^* \cup \{1\}^* \]

\[ a^{-1} \cdot B = \{\alpha \in \Sigma^* \mid a\alpha \in B\} \]

**Example**

\[ 0^{-1}A = \{1\}^* = 1^{-1}A \]

\[ (0^2)^{-1}A = \emptyset, \text{ etc.} \]

\[ Q_A = \{A, 0^{-1}A, \emptyset\} \]

\[ T_A = \{A, 0^{-1}A\} \]
Theorem

Let $A \subseteq \Sigma^*$ be recognizable and suppose that $\mathcal{M} = (\mathcal{M}, i, T)$, where $\mathcal{M} = (Q, \Sigma, F)$, is a complete accessible recognizer with behaviour $A$. There exists a function $f: Q \rightarrow Q_A$ such that:

(i) $f(i) = i_A$,

(ii) $f^{-1}(T_A) = T$,

(iii) $(f(q))F^{A}_\sigma = f(qF_\sigma)$ for all $q \in Q$, $\sigma \in \Sigma$,

(iv) $f$ is surjective.
Theorem

Let $A \subseteq \Sigma^*$, then $A$ is recognizable if and only if the collection \( \{ \beta^{-1}A \mid \beta \in \Sigma^* \} \) is finite.
Recognizable sets

\(A, B\) are recognizable subsets of \(\Sigma^*\)

Let \(A, B \subseteq \Sigma^*\). If \(A\) and \(B\) are recognizable then \(A \cup B\) is also recognizable.

Let \(A, B \subseteq \Sigma^*\). If \(A\) and \(B\) are recognizable sets then \(A \cdot B\) is also recognizable.

Let \(A \subseteq \Sigma^*\). If \(A\) is recognizable then so is \(A^*\).

Let \(A \subseteq \Sigma^*\) be a recognizable set, then \(\Sigma^* \setminus A\) is also recognizable.

If \(A\) and \(B\) are recognizable subsets of \(\Sigma^*\) then so is \(A \cap B\).

Let \(\Sigma, \Gamma\) be finite non-empty sets and \(f: \Sigma^* \to \Gamma^*\) a function satisfying the condition \(f^{-1}(\Lambda_\Gamma) = \Lambda_\Sigma\) where \(\Lambda_\Gamma\) and \(\Lambda_\Sigma\) are the empty words in \(\Gamma^*\) and \(\Sigma^*\) respectively. If \(A \subseteq \Sigma^*\) is recognizable then so is \(f(A)\).
Congruence relation

\[ \mathcal{M} = (Q, \Sigma, F) \quad \alpha \sim_\mathcal{M} \beta \iff F_\alpha = F_\beta \quad \sim_\mathcal{M} \text{ is a congruence on } \Sigma^* \]

\[ \mathcal{N} = (\mathcal{M}, i, T) \]
\[ \alpha \sim_\mathcal{M} \beta \iff [x\alpha y \in |\mathcal{N}| \iff x\beta y \in |\mathcal{N}|, \text{ for all } x, y \in \Sigma^*]. \]

\[ A \subseteq \Sigma^* \]
\[ \alpha \cong_A \beta \iff [x\alpha y \in A \iff x\beta y \in A \text{ for all } x, y \in \Sigma^*]. \]
\[ \alpha \sim_\mathcal{M} \beta \iff \alpha \cong_{|\mathcal{N}|} \beta. \]
Theorem

Let $A \subseteq \Sigma^*$ be a recognizable subset of $\Sigma^*$ with minimal complete recognizer $\mathcal{M}_A = (M_A, i_A, T_A)$. Then for $\alpha, \beta \in \Sigma^*$ we have

$$\alpha \sim_{M_A} \beta \iff \alpha \simeq_A \beta.$$
Syntactic monoid

recognizable set \( A \subseteq \Sigma^* \)

congruence \( \approx_A \) \hspace{2cm} \text{Myhill congruence of } A

\[ \Sigma^*/\approx_A = \Sigma^*/\sim_{M_A} \cong M(M_A) \]

\text{syntactic monoid of } A
Let $A \subseteq \Sigma^*$. The following statements are equivalent:

(i) $A$ is recognizable.

(ii) $\Sigma^*/\approx_A$ is finite.

(iii) $A$ is the union of congruence classes of a congruence on $\Sigma^*$ of finite index.
$A = \{0\}^* \cdot \{1\} \cdot \{0\}^*$

Example

$\{\Lambda, 1, 1^2\}$

\[
\begin{array}{c|ccc}
\Lambda & 1 & 1^2 \\
\hline
\Lambda & 1 & 1^2 & 1^2 \\
1 & 1 & 1^2 & 1^2 \\
1^2 & 1^2 & 1^2 & 1^2 \\
\end{array}
\]
Rational operations

\[ \Sigma^* \]

set \( \mathcal{P}(\Sigma^*) \) consisting of all sets of words in \( \Sigma^* \).

\[ A \cup B \]

\[ A \cdot B \]

\[ A^* \]

**rational operations** on \( \mathcal{P}(\Sigma^*) \) where \( A, B \in \mathcal{P}(\Sigma^*) \)

Now let \( \mathcal{H} \subseteq \mathcal{P}(\Sigma^*) \), we say that \( \mathcal{H} \) is **closed under the rational operations** if given \( A, B \in \mathcal{H} \) then \( A \cup B \in \mathcal{H}, A \cdot B \in \mathcal{H} \) and \( A^* \in \mathcal{H} \).
\[ \text{Rat}(\Sigma) \subseteq \mathcal{P}(\Sigma^*) \]

is the smallest subset of \( \mathcal{P}(\Sigma^*) \)
contains the singleton subsets and \( \emptyset \)
is closed under the rational operations.
Regular set

set \( A \in \mathcal{P}(\Sigma^*) \)

\[ \emptyset \quad \{ x \} \text{ (where } x \in \Sigma^*) \]

formed from sets of this type

described as a finite number of rational operations

\( A \in \text{Rat}(\Sigma) \).

\[ \downarrow \]

regular set

collection of all regular words

\( \text{Reg}(\Sigma) = \text{Rat}(\Sigma) \)
Proposition

Let $\Sigma$ and $\Gamma$ be non-empty finite sets and suppose that $f : \Sigma \to \Gamma$ is a mapping. Define $f^* : \Sigma^* \to \Gamma^*$ by

$$f^*(\sigma_1 \ldots \sigma_n) = f(\sigma_1) \ldots f(\sigma_n), \quad \sigma_1 \ldots \sigma_n \in \Sigma^+$$

$$f^*(\Lambda) = \Lambda.$$

If $A$ is a regular set of $\Sigma^*$ then $f^*(A)$ is a regular set of $\Gamma^*$. 
$S$ be any finite non-empty set  \( R \) is a relation
\[ aRa' \] to mean \((a, a') \in R\) or \(a\) is related to \(a'\) under \(R\).
\[ \alpha = s_1 \ldots s_n \in S^{+} \] \(R\)-word if \(s_i R s_{i+1}\) for all \(i = 1, \ldots, n - 1\).
\[ \alpha = s_1 \ldots s_n \text{ and } \alpha' = s'_1 \ldots s'_n \]
\[ \alpha \cdot \alpha' = s_1 \ldots s_n \cdot s'_1 \ldots s'_n \text{ if } s_n R s'_1. \]
\[ \text{two } R\text{-words } \alpha = s_1 \ldots s_n \text{ and } \alpha' = s'_1 \ldots s'_n \]
\[ \alpha \cdot \alpha' = s_1 \ldots s_n \cdot s'_1 \ldots s'_n \text{ if } s_n R s'_1. \]
\[ \text{two sets } X, Y \text{ of } R\text{-words then we define the sets} \]
\[ X \cup Y \]
\[ X \cdot Y = \{x \cdot y \mid x \in X, y \in Y \text{ and } x \cdot y \text{ is an } R\text{-word} \} \]
\[ X^{*} = \{x_1 \cdot x_2 \ldots x_m \mid x_i \in X \text{ and } x_1 \cdot x_2 \ldots x_m \text{ is an } R\text{-word} \}. \]
\[ \text{all sets of } R\text{-words in } S^{*}. \]

Given \(s_1, s_n \in S\) we define \(R(s_1, s_n)\) to be the set of all \(R\)-words in \(S^{*}\)
of the form \(s_1 \ldots s_n\).

Let \(S\) be a non-empty finite set, \(R\) a binary relation on \(S\) and \(s_1, s_n \in S\), then the set \(R(s_1, s_n)\) is a regular set of words of \(S^{*}\).
Theorem

If \( A \) is a recognizable set of \( \Sigma^* \) then \( A \) is regular.

(Kleene) Let \( \Sigma \) be a finite non-empty set. The class of recognizable sets of \( \Sigma^* \) equals the class \( \text{Reg}(\Sigma) \) of all regular sets of \( \Sigma^* \).
Direct recognizer

Let $\mathcal{M} = (M, i, T)$ be a recognizer such that $T$ is a singleton; we call $\mathcal{M}$ a **direct recognizer**.

A recognizable set $A \subseteq \Sigma^*$ is called **unitary** if the minimal complete recognizer $\mathcal{M}_A$ is direct.
Theorem

Let \( A \subseteq \Sigma^* \) be a recognizable set. Then \( A \) is unitary if and only if \( A \neq \emptyset \) and \( \alpha^{-1} \cdot A = \beta^{-1} \cdot A \) for all \( \alpha, \beta \in A \).

Let \( A \subseteq \Sigma^* \) be recognizable with \( A \neq \emptyset \), then \( A = \bigcup_{j=1}^{r} A_j \) where the \( A_j \) are unitary and \( A_j \cap A_k = \emptyset \) if \( j \neq k \).

We call the sets \( A_j \ (j = 1, \ldots, r) \) the unitary components of \( A \).
Example

\[ A = \{0\} \cdot \{10\}^* \cup \{01\}^+ \]

\[ A = B_1 \cup B_2 \cup B_3 \]

\[ B_1 = \{0\}; \quad B_2 = \{010\} \cdot \{10\}^*; \quad B_3 = \{01\}^+ \]

unitary sets

\[ A = A_1 \cup A_2 \]

\[ A_1 = \{0\} \cdot \{10\}^* \quad A_2 = \{01\}^+ \]

unitary sets
\( \hat{A} \subseteq \Sigma^* \)

\[
\alpha^{-1} \cdot A = \{ \Lambda \} \text{ for all } \alpha \in A
\]

**Prefix**

The set \( A \) is a prefix if there are no words of the form \( \alpha = \beta \gamma \) where both \( \alpha \) and \( \beta \) belong to \( A \).

Let \( A \subseteq \Sigma^* \) be a recognizable set. Then \( A \) is a prefix if and only if the minimal complete recognizer \( \mathcal{M}_A \) is direct and \( T_A \ast \Sigma = \emptyset \).

**Prefix part of \( A \)**

\[
A_p = A \backslash A \cdot \Sigma^+ 
\]
Unitary monoid

Let $B \subseteq \Sigma^*$; we call $B$ a **unitary monoid** if

(i) $B$ is a unitary subset of $\Sigma^*$ ($B$ is thus recognizable);
(ii) $B$ is a submonoid of $\Sigma^*$.

Our basic aim is the decomposition of a recognizable set into subsets of the form $A \cdot M$ where $A$ is a prefix and $M$ is a unitary monoid.
Theorem

Given any unitary subset $A$ of $\Sigma^*$ the set

$$A_M = A^{-1} \cdot A = \{ \gamma \in \Sigma^* | \alpha \gamma \in A \text{ for some } \alpha \in A \}$$

is a unitary monoid and $A = A_P \cdot A_M$.

Let $A$ be a recognizable subset of $\Sigma^*$. Then

$$A = B_1 C_1 \cup B_2 C_2 \cup \ldots \cup B_r C_r$$

where $B_i C_i$ are unitary subsets, $B_i$ are prefixes and $C_i$ are unitary monoids for $i = 1, 2, \ldots, r$.

unitary-prefix decomposition.
Example

\[ A = \{0\}^* \cdot \{\{10\} \cup \{0\}\}^* \cdot \{0\}^* \]

\[ 0^{-1} \cdot A = A \]
\[ 1^{-1} \cdot A = 0B \]
\[ (01)^{-1} \cdot A = 1^{-1} \cdot A = 0B \]
\[ (10)^{-1} \cdot A = B \]
\[ (11)^{-1} \cdot A = \emptyset \]
\[ (100)^{-1} \cdot A = B \]
\[ (101)^{-1} \cdot A = 0B \]

\[ A = \{0\}^* \cup \{0\}^*\{10\}B \]
Pumping lemma

(Pumping lemma) Let $A \subseteq \Sigma^*$ be recognizable and suppose that $n = |Q_A|$, the number of states in the minimal complete recognizer of $A$. If $\alpha \in A$ and the length of $\alpha$ is greater than or equal to $n$ then

$$\alpha = \beta \gamma \delta$$

such that

(i) $\gamma \neq \Lambda$,
(ii) $\{\beta\} \cdot \{\gamma\}^* \cdot \{\delta\} \subseteq A$. 
Example

\[ A = \{0\}^* \cdot \{\{10\} \cup \{0\}\}^* \cdot \{0\}^* \]

\[ n = 4 \]

\[ \alpha = 00010100 \]
\[ \beta = 000, \ \gamma = 1010, \ \delta = 0 \quad \alpha = \beta \gamma \delta \]
\[ \beta' = 0, \ \gamma' = 00, \ \delta' = 10100 \]
define $A^{(n)}$ to be the set of all words of $A$ that are of length $n$ for $n = 0, 1, \ldots$

$$A = \bigcup_{n=0}^{\infty} A^{(n)}.$$ 

For a finite set $A$ we will have $A^{(n)} = A^{(n+1)} = \ldots = \emptyset$. For an infinite set each $A^{(n)}$ is finite, in fact $|A^{(n)}| \leq k^n$ where $k = |\Sigma|$. 
let $\mathcal{M} = (Q, \Sigma, F)$ be a complete finite state machine $|Q| = m$.

Set of $m \times m$ matrices

First we let $Q = \{q_1, \ldots, q_m\}$ and then for each $\sigma \in \Sigma$ define the matrix

$$f_\sigma = (f_{ij}^\sigma) \quad \text{where} \quad f_{ij}^\sigma = \begin{cases} 1 & \text{if } q_iF = q_j \\ 0 & \text{otherwise} \end{cases} \quad \text{for } i, j \in \{1, \ldots, m\}.$$ 

Each row of the matrix $f_\sigma$ will consist of one 1 and $(m - 1)$ 0s.

Each state $q_i$ will be represented by a $1 \times m$ row vector $e_j$ of the form $(0 \ldots 010 \ldots 0)$ with a 1 in the $j$-th position.

$$q_iF_\sigma = q_k \iff e_j \cdot f_\sigma = e_k.$$ 

Given $\alpha = \sigma_1 \ldots \sigma_n \in \Sigma^*$ we define $f_\alpha = f_{\sigma_1} \cdots f_{\sigma_n}$

$$q_iF_\alpha = q_k \iff e_j \cdot f_\alpha = e_k.$$ 

Put $f_\Lambda = I_m$, the $m \times m$ identity matrix.
If $\mathcal{M}=(M, i, T)$ is a recognizer, let $i=q_1$ and define
\[ \mathcal{E} = \{ e_j | q_j \in T \}, \]
then for each $\alpha \in |\mathcal{M}|$ we have $e_1 \cdot f_\alpha \in \mathcal{E}$ and clearly
\[ |\mathcal{M}| = \{ \alpha \in \Sigma^* | e_1 \cdot f_\alpha \in \mathcal{E} \}. \]

Let $F = \sum_{\sigma \in \Sigma} f_\sigma$, which is again an $m \times m$ matrix (it belongs to the set of all $m \times m$ matrices over the integers); we call $F$ the matrix of $\mathcal{M}$.

For any subset $R \subseteq Q$ we define
\[ \mathcal{E}(R) = \{ e_j | q_j \in R \} \]
and consider
\[ \mathcal{E}(R) = \sum_{e_j \in \mathcal{E}(R)} e_j^T. \]

($e_j^T$ is the transpose of $e_j$ and thus $\mathcal{E}(R)$ is a column vector.)
Theorem

Let $\mathcal{M} = (M, i, T)$ be a recognizer with matrix $\mathcal{F}$. Let $R$ be a set of states of $\mathcal{M}$ and $k \geq 0$, then the number of words of $\Sigma^*$ of length $k$ which send the initial state $i$ to a state in $R$ is given by

$$e_1 \cdot (\mathcal{F}^k) \cdot \mathcal{G}(R).$$

The number of words in $|\mathcal{M}|$ of length $k$ is given by

$$e_1 \cdot (\mathcal{F}^k) \cdot \mathcal{G}(T).$$

The total number of words in $|\mathcal{M}|$ is given by

$$e_1 \cdot (\mathcal{F}^0 + \mathcal{F}^1 + \mathcal{F}^2 + \ldots + \mathcal{F}^k + \ldots) \cdot \mathcal{G}(T) = e_1 \cdot (I - \mathcal{F})^{-1} \cdot \mathcal{G}(T).$$
Example

The number of words of $|\mathcal{M}|$ of length 2 is given by

$$\mathcal{E}_1 \cdot \mathcal{F}^2 \cdot \mathcal{C}(T) = (1, 0, 0) \cdot \begin{pmatrix} 0 & 1 & 3 \\ 0 & 1 & 3 \\ 0 & 0 & 4 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = 1.$$ 

$$\mathcal{E}_0 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathcal{E}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix}, \quad \mathcal{F} = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & 2 \end{pmatrix},$$

$$\mathcal{E}_1 = (1, 0, 0), \quad \mathcal{C}(T) = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}.$$