I.5.1. Regular Grammars and NFAs (13.5)

We will show that regular expressions coincide with regular languages and with languages recognised by a DFA or NFA. Here we prove one part of this result:

Theorem (I.5.1.1.)

For every right linear grammar $G$ there exists an NFA $A$ s.t.

$$L(G) = L(A)$$

A can be computed from $G$. 
Proof Idea

A derivation of a word in $G$ has the form

$$S = A_0 \rightarrow a_1 A_1 \rightarrow a_1 a_2 A_2 \rightarrow \cdots \rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1}$$

$$\rightarrow a_1 a_2 \cdots a_{n-1} a_n$$

where we have productions

$$A_i \rightarrow a_{i+1} A_{i+1} \quad A_{n-1} \rightarrow a_n$$

or

$$S = A_0 \rightarrow a_1 A_1 \rightarrow a_1 a_2 A_2 \rightarrow \cdots \rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1}$$

$$\rightarrow a_1 a_2 \cdots a_{n-1}$$

where we have productions

$$A_i \rightarrow a_{i+1} A_{i+1} \quad A_{n-1} \rightarrow \varepsilon$$

Proof Idea

Define $A$ with states $N \cup \{q_F\}$ for a special new accepting state $q_F$ s.t. the derivation

$$S = A_0 \rightarrow a_1 A_1 \rightarrow a_1 a_2 A_2 \rightarrow \cdots \rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1}$$

$$\rightarrow a_1 a_2 \cdots a_{n-1} a_n$$

corresponds to a sequence of transitions

$$S = A_0 \xrightarrow{a_1} A_1 \xrightarrow{a_2} A_2 \xrightarrow{a_3} \cdots \xrightarrow{a_{n-1}} A_{n-1} \xrightarrow{a_n} q_F$$

and a derivation

$$S = A_0 \rightarrow a_1 A_1 \rightarrow a_1 a_2 A_2 \rightarrow \cdots \rightarrow a_1 a_2 \cdots a_{n-1} A_{n-1}$$

$$\rightarrow a_1 a_2 \cdots a_{n-1}$$

corresponds to a sequence of transitions

$$S = A_0 \xrightarrow{a_1} A_1 \xrightarrow{a_2} A_2 \xrightarrow{a_3} \cdots \xrightarrow{a_{n-1}} A_{n-1} \in F$$

Constructed NFA

We obtain from $G = (N, T, S, P)$ the following NFA:

<table>
<thead>
<tr>
<th>automaton</th>
<th>$A$</th>
</tr>
</thead>
<tbody>
<tr>
<td>states</td>
<td>$N \cup {q_F}$</td>
</tr>
<tr>
<td>terminals</td>
<td>$T$</td>
</tr>
<tr>
<td>start</td>
<td>$S$</td>
</tr>
<tr>
<td>final</td>
<td>$B \in N$ s.t. $B \rightarrow \varepsilon.$</td>
</tr>
<tr>
<td></td>
<td>$q_F$</td>
</tr>
</tbody>
</table>
| transitions      | $B \xrightarrow{a} B'$ if $B \rightarrow aB'$.
|                  | $B \xrightarrow{a} q_F$ if $B \rightarrow a$. |
Proof of Theorem I.5.1.1.

Can be found in the additional material.

**Example**

Consider the Grammar:

\[
\text{grammar } G \\
\text{terminals } 0, 1 \\
\text{nonterminals } S, T \\
\text{start symbol } S \\
\text{productions } S \rightarrow 0, S \rightarrow 1 T, \\
\quad T \rightarrow 0 T, T \rightarrow 1 T, \\
\quad T \rightarrow \epsilon, T \rightarrow 0, T \rightarrow 1
\]

**Corresponding Automaton**

(Note that it is nondeterministic).

With corresponding rules:

- \( S \rightarrow 0 \)
- \( S \rightarrow 1 T \)
- \( T \rightarrow 0 T \)
- \( T \rightarrow 1 T \)
- \( T \rightarrow \epsilon \)
- \( T \rightarrow 0 \)
- \( T \rightarrow 1 \)

Accepting state because of \( T \rightarrow \epsilon \)
Theorem I.5.2.1.

Let $A = (Q, q_0, F, T, \rightarrow)$ be an NFA. Then there exist a regular expression $E$ s.t. $L(E) = L(A)$. $E$ can be computed from $A$.

Proof of Theorem I.5.2.1.

A proof of Theorem I.5.2.1. and an example can be found in the additional material.
I.5.3. Equivalence Theorem

Theorem I.5.3.1.

Theorem (I.5.3.1.)

Let $L$ be a language over an alphabet $T$. The following are equivalent:

1. $L$ is definable by a regular expression.
2. $L$ is a regular.
3. $L$ is definable by a right-linear grammar.
4. $L$ is definable by a left-linear grammar.
5. $L$ is definable by an NFA with empty moves
6. $L$ is definable by an NFA.
7. $L$ is definable by a DFA.

Furthermore, the corresponding regular expressions, right linear grammars, left-linear grammars, NFAs with empty moves, NFAs, DFAs can be computed from each other.

Proof of Theorem I.5.3.1.

A proof of Theorem I.5.3.1. can be found in the additional material.

---

I.5.4. Closure Properties and Decidability of Regular Languages

Closure Properties

Theorem (I.5.4.1.)

Regular languages are closed under
1. complement,
2. intersection,
3. the operation $L \mapsto L^R$.

So if $L, L'$ are regular languages over alphabet $T$, so are
1. $L^c$ (the complement of $L$, i.e. $\{t \in T^* \mid t \notin L\}$),
2. $L \cap L'$,
3. $L^R$ (i.e. $\{w^R \mid w \in L\}$, where $w^R$ is the result of rewerting $w$).

Furthermore, regular grammars for $L^c$, $L \cap L'$ and $L^R$ can be computed from those for $L$. 
Proofidea for Theorem I.5.4.1.

- We will use that regular expressions, languages definable by DFAs and regular languages are equivalent.
- Languages definable by regular expressions are closed under $L \mapsto L^R$. (Lemma I.5.3.4).
- We will show that languages definable by a DFA are closed under $L \mapsto L^c$ and under intersection.
- Therefore the assertion follows.
- Full details can be found in the additional material.

Decision Problems

Theorem (I.5.4.3.)

- We can decide for regular languages whether $L = \emptyset$.
- We can decide for regular languages $L$ and $L'$ whether $L \subseteq L'$.
- We can decide for regular languages $L$ and $L'$ whether $L = L'$.

Proof of Theorem I.5.4.3.

A proof of Theorem I.5.4.3. can be found in the additional material.
Motivation

- We want to show that there are languages which are context-free but not regular.
- In order to do this we prove the pumping lemma, which uses the fact that an NFA has only finitely many states. (We could use as well the fact that a regular grammar has only finitely many nonterminals).
- **Note** The following slides contain some coloured parts. The colours are indistinguishable in the black and white handouts. It is recommended to look at them using the online version.

Using the Finiteness of an NFA

Consider an NFA

![Diagram of an NFA](image)

This NFA has 5 states. Any run of the NFA for a word of length \( \geq 5 \) uses at least 6 states. Therefore it must visit one state at least twice. So there must be a loop within the first 5 letters of such a word.

---

Using the Finiteness of an NFA

Here is the run for the word \( z = ababa \) using colours **blue**, **red** and **green**.

![Diagram of NFA run](image)

- The **blue part** is the part before we reached a state visited twice, corresponding to the word \( u = a \).
- The **red part** is the part from the state visited twice until we reach it again, corresponding to the word \( v = bab \).
- The **green part** is the remaining part, corresponding to the word \( w = a \).
- The loop must occur within the first 5 letters, so \( |uv| \leq 5 \). Because \( v \) is along a loop, \( |v| \geq 1 \).

---

Using the Finiteness of an NFA

If we repeat the loop several times, we obtain as well an accepting word.

- If we start with \( u = a \), then repeat the loop following the word \( v = bab \) \( i \) times, then the follow the word \( w = a \), we obtain an accepting run.
- It accepts the word \( a(bab)^ia \).
  - E.g. in case \( i = 2 \) the word is \( ababbaba \).
  - In case \( i = 0 \) the word is \( aa \).
- In general we get that the word \( uv^iw \) is an element of the language as well.
I.5.5. The Pumping Lemma for Regular Languages (12.4, 12.5)

Generalisation

Assume an NFA $A$ having $k$ states. Then for every word $x \in L(A)$ s.t. $|x| \geq k$ there exist words $u, v, w$ s.t.

$x = uvw$, $|uv| \leq k$, $|v| \geq 1$

and s.t.

$uv^i w \in L(A)$ for all $i \in \mathbb{N}$

This follows by the above considerations.

So we have proved the following theorem:

Example 1

Lemma

The language $L := \{a^ib^i \mid i \geq 1\}$ is context-free but not regular.

Proof (Example 1)

- We have already seen that $L$ is context-free.
- Assume $L$ is regular.
- Let $k$ be as in the pumping lemma.
- Consider $x := a^kb^k \in L$.
- $|x| \geq k$, so there exist $u, v, w$ s.t.

  $x = uvw$, $|uv| \leq k$, $|v| \geq 1$,

  and s.t.

  $uv^i w \in L$ for all $i \in \mathbb{N}$.

- Since $|uv| \leq k$, $u$ and $v$ are substrings of $a^k$.
- Therefore $uv^2w = a^{k+l}b^k$ where $l = |v|$.
- But $a^{k+l}b^k \notin L$, a contradiction.
The language $L := \{xx^R \mid x \in \{a, b\}^*\}$ is context-free but not regular.

Proof (Example 2)

- We have already seen that $L$ is context-free.
- Assume $L$ is regular.
- Let $k$ be as in the pumping lemma.
- Consider $x := a^kbb^a^k \in L$.
- $|x| \geq k$, so there exist $u, v, w$ s.t.
  - $x = uvw$, $|uv| \leq k$, $|v| \geq 1$,
  - and s.t. $uv^i w \in L$ for all $i \in \mathbb{N}$.
- Since $|uv| \leq k$, $u$ and $v$ are substrings of $a^k$.
- Therefore $uv^{2}w = a^{k+l}bb^a^k$ where $l = |v|$.
- But $a^{k+l}bb^a^k \not\in L$, a contradiction.