The CORRECTNESS PROBLEM

Q: How can we justify the CFG we designed?

For instance, let us consider the following CFG

$$S \to aB \mid bA$$
$$A \to aS \mid bAA \mid a$$
$$B \to bS \mid aBB \mid b$$

and prove that

claim → $L(G) = \{ w \in \{a,b\}^+ \mid \#_a(w) = \#_b(w) \}$.

$$L_a = \{ w \in \{a,b\}^+ \mid \#_a(w) = \#_b(w) + 1 \}$$

$$L_b = \{ w \in \{a,b\}^+ \mid \#_b(w) = \#_a(w) + 1 \}$$

$$L_= = L(G)$$

To prove the claim above, we need to prove the following:

(1) $A \Rightarrow^* w$ iff $w \in L_a$

(2) $B \Rightarrow^* w$ iff $w \in L_b$

(3) $S \Rightarrow^* w$ iff $w \in L_=$

Proof: "$\Rightarrow$" (the only-if direction)

by induction on the length of derivations:

Basis:

1) $A \Rightarrow^* w$ length 1.

Then, we have $A \Rightarrow a \in L_a$

2) $B \Rightarrow^* w$ with length 1.

$\Rightarrow B \Rightarrow b \in L_b$

3) $S \Rightarrow^* w$ with length 2. (Why?)

Thus, it must be either

- $S \Rightarrow aB \Rightarrow ab \in L_=$

or

- $S \Rightarrow bA \Rightarrow ba \in L_=$.

Inductive Step:

Suppose for some odd integer $m \geq 1$ that

Note: This is inductive hyp.

(1) $A \xRightarrow{n} w \Rightarrow w \in L_a$ (length of derivation)

(2) $B \xRightarrow{n} w \Rightarrow w \in L_b$

(3) $S \xRightarrow{n+1} w \Rightarrow w \in L_=$.

for all odd n s.t. $1 \leq n \leq m$.

For part (1): Consider a longer derivation $A \xRightarrow{n+2} w$. Note that there are two possible first steps:

Case 1.1: $A \Rightarrow aS \xRightarrow{n+1} a \cdot y = w$

Since $S \xRightarrow{n+1} y$ is of length $m+1$, by ind. hyp. $y \in L_=$.

$w = a \cdot y \in L_a$.

Case 1.2:

$A \Rightarrow bAA \xRightarrow{n_1} b x A \xRightarrow{n_2} b x y = w$

Since $A \xRightarrow{n_1} x$, $A \xRightarrow{n_2} y$ and $n_1, n_2 \leq m$

by ind. hyp. $x, y \in L_a$

Recall the CFG:

$S \rightarrow aB \mid bA$

$A \rightarrow aS \mid bAA \mid a$

$B \rightarrow bS \mid aBB \mid b$

Note that $w = b\,x\,y \in L_a$

because $\#_a(w) = \#_b(w) + 1$.

For part (2): Consider a longer derivation $B \overset{m+2}{\Longrightarrow} w$.

As in case (1), by using similar arguments, we obtain $w \in L_b$.

e.g. $B \Longrightarrow bS \overset{m+1}{\Longrightarrow} by \ldots y \in L_=$, $by \in L_b$.

For part (3): Consider a longer derivation $S \overset{m+3}{\Longrightarrow} w$.

For the first step there are two possibilities

Case 3.1: $S \Longrightarrow aB \overset{m+2}{\Longrightarrow} ay = w$.

As proved in (2) $B \overset{m+2}{\Longrightarrow} y$ implies $y \in L_b$.

Then, $w = ay \in L_=$.

Case 3.2: $S \Longrightarrow bA \overset{m+2}{\Longrightarrow} bx = w$

By similar argument as in case 3.1, we have $x \in L_a$, $w = bx \in L_=$.

This concludes the inductive step, and hence "$\Longrightarrow$"-direction is proven.

We now prove "$\Longleftarrow$" - direction

We show that

(1) $w \in L_a$ implies $A \overset{+}{\Longrightarrow} w$

(2) $w \in L_b$ implies $B \overset{+}{\Longrightarrow} w$

(3) $w \in L_=$ implies $S \overset{+}{\Longrightarrow} w$.

Proof: By induction on $|w|$.

Basis:

(1) $w = a \Longrightarrow A \Longrightarrow a$

(2) $w = b \Longrightarrow B \Longrightarrow b$

(3) $w = ab$ or $w = ba$

Clearly, $S \Longrightarrow aB \Longrightarrow ab$

and $S \Longrightarrow bA \Longrightarrow ba$.

Inductive Step:

Inductive Hypothesis:

Suppose for some odd $m \geq 1$ that

(1) $w \in L_a \wedge |w| = n$ implies $A \overset{*}{\Longrightarrow} w$

(2) $w \in L_b \wedge |w| = n$ implies $B \overset{*}{\Longrightarrow} w$

(3) $w \in L_= \wedge |w| = n+1$ implies $S \overset{*}{\Longrightarrow} w$

for all odd n s.t. $1 \leq n \leq m$.

For part (1): Let $w \in L_a$ and $|w| = m+2$.

Note that w begins with a or b.

Case (1.1): $w = bx$.

Then $|x| = m+1$. and

$$\#_a(x) = \#_b(x) + 2.$$

Q: How to generate bx from A.

It has to be of the form

$$A \Longrightarrow bAA \overset{*}{\Longrightarrow} bx.$$

i.e. $AA \overset{*}{\Longrightarrow} yA \overset{*}{\Longrightarrow} yz = x.$

Recall:

$S \to aB \mid bA$
$A \to aS \mid bAA \mid a$
$B \to bS \mid aBB \mid b$

If this is to be achieved, then one should be able to write $x = yz$ and $y, z \in L_a$.

Indeed one must be able to prove the following fact:

Fact: If $\#_a(x) = \#_b(x) + 2$, then x can be written as $x = yz$. s.t. $y, z \in L_a$.

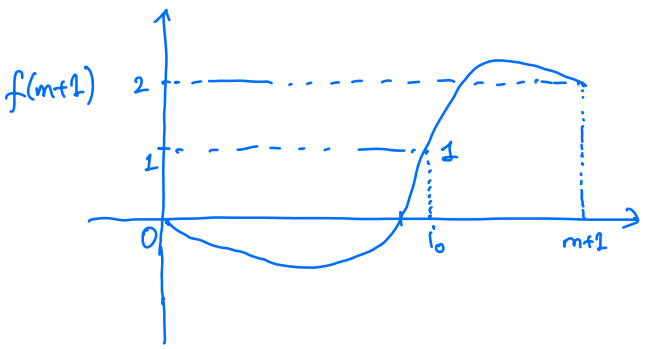
Proof of the fact: Let $x = x_1 \ldots x_{m+1}$.

(Recall $|x| = m+1$).

Define a function $f: \{0, 1, \ldots, m+1\} \to \mathbb{Z}$.

by $f(i) = \#_a(x_1 \ldots x_i) - \#_b(x_1 \ldots x_i)$.

Then $f(0) = 0$ by definition and
$f(m+1) = 2$ by assumption.
The graph of f has the form



So there exists $0 < i_0 < m+1$ s.t.
$$f(i_0) = 1.$$
Letting $y = x_1 \dots x_{i_0}$ and $z = x_{i_0+1} \dots x_{m+1}$,
we get $y, z \in L_a \wedge x = yz$. □

Now, both $y, z \in L_a$ and $|y|, |z| \leq m$,
so by inductive hypothesis
$$A \Rightarrow^* y, \quad A \Rightarrow^* z.$$
Therefore, we have $A \Rightarrow bAA \Rightarrow^* byA \Rightarrow^* byz = w$.

Case (1.2): $w = ax$.

Then, $x \in L_=$ and $|x| = m+1$.
By inductive hypothesis, $S \Rightarrow^* x$.
Therefore, $A \Rightarrow aS \Rightarrow^* ax = w$.

For part (2): Similar to (1).

Exercise: Try working out the details for part (2) on your own.

For part (3): Let $w \in L_=$ and $|w| = m+3$.

Case (3.1) $w = ax$.

Then $x \in L_b$, $|x| = m+2$.

From (2) we get that $B \Rightarrow^* x$.

Therefore, $S \Rightarrow aB \Rightarrow^* ax = w$.

Case (3.2): $w = bx$.

Then $x \in L_a$, $|x| = m+2$.

From (1), it follows that $A \Rightarrow^* x$.

Then, $S \Rightarrow bA \Rightarrow^* bx = w$.

This completes the proof of "$\Leftarrow$"-direction.

Since we proved both $\Rightarrow$ and $\Leftarrow$ directions, we are done with proof of the claim. $\square$

REGULAR GRAMMARS

A subclass of CFGs that generate regular languages.

Definition: A reg. grammar is a CFG in which every rule is in one of the following forms:

- $A \to aB$
- $A \to a$
- $A \to \varepsilon$

for all $A, B \in V$ and $a \in \Sigma$.

Example:

$S \to aB \mid \varepsilon$

$B \to bS \mid bA$

$A \to aA \mid \varepsilon$



$L(G) = (ab)^+ a^* + \varepsilon$ (Why?)

Example:

$S \to aB \mid bA$

$A \to aS \mid bAA \mid a$

$B \to bS \mid aBB \mid b$

Not regular! $A \to bAA$ is not in correct form.

Regular Grammars & Finite Automata

Goal: To show reg. grammars $\Leftrightarrow$ FA.

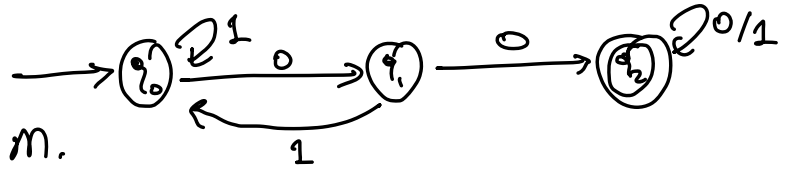
Proposition: $L = L(M)$ for some DFA M
$\Rightarrow$ $L = L(G)$ for some reg. gr. G.

Proof: Given a DFA $M = (Q, \Sigma, \delta, q_0, F)$, the reg. grammar $G = (V, \Sigma, R, S)$ is constructed as follows

- $V = Q$ $\qquad \Sigma_M = \Sigma_G$
- $S = q_0$
- R contains the following rules
  - $\delta(q, a) = q' \iff q \longrightarrow aq'$
  - For each $q_f \in F$, add $q_f \to \varepsilon$.

- Transitions $\longrightsquigarrow$ Productions
- Computation $\longrightsquigarrow$ Derivations

Example:



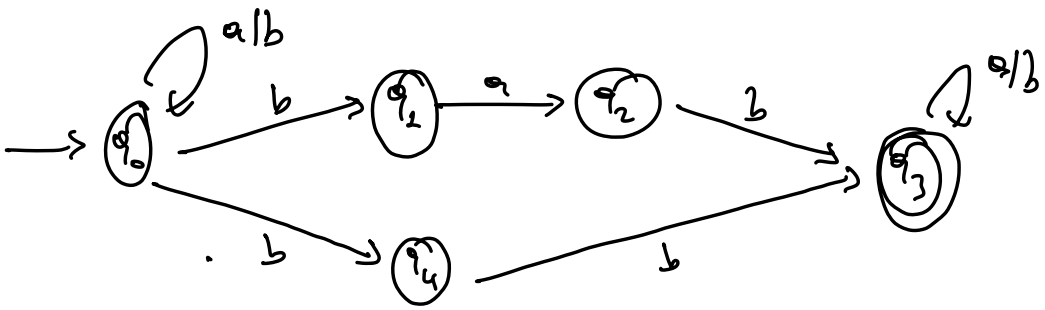
DFA M.

CFG G:

$q_0 \to 1q_0$
$q_0 \to 0q_1$
$q_1 \to 1q_0$
$q_1 \to 0q_2$
$q_2 \to 1q_2$
$q_2 \to 0q_2$
$q_2 \to \varepsilon$.

$q_0 \to 1q_0 \mid 0q_1$
$q_1 \to 1q_0 \mid 0q_2$
$q_2 \to 1q_2 \mid 0q_2 \mid \varepsilon$

Example: Same technique applies to NFAs as well.



$q_0 \to aq_0 \mid bq_0 \mid bq_1 \mid bq_4$
$q_1 \to aq_2$
$q_2 \to bq_3$
$q_4 \to bq_3$
$q_3 \to aq_3 \mid bq_3 \mid \varepsilon$

Proposition: $L = L(G)$ for some reg. grammar G.
$\Rightarrow$ $L = L(M)$ for some NFA M.

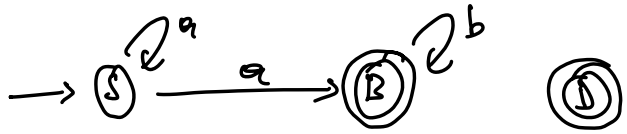
Proof: For a given reg. grammar $G = (V, \Sigma, R, S)$, the equivalent NFA $N = (Q, \Sigma, \delta, q_0, F)$ is constructed as follows:

1) $Q = V \cup \{D\}$
2) $q_0 = S$
3) $F = \{A \in V \mid A \to \varepsilon \in R\} \cup \{D\}$
4) δ is defined by
   - $\delta(A, a) = B$ iff $A \to aB \in R$.
   - $\delta(A, a) = D$ iff $A \to a \in R$

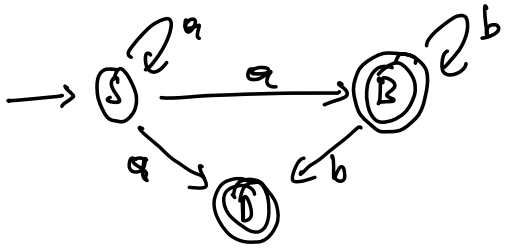
□

Example: G: $S \to aS \mid aB$
$B \to bB \mid \varepsilon$

The equivalent NFA N is



* Suppose G also contains $S \to a$ and $B \to b$



Example:

$S \rightarrow 0S \mid 1S \mid 0A$

$A \rightarrow 0B \mid 1$

$B \rightarrow 0B \mid 1B \mid \varepsilon$

