

4. Context-free Languages

4.1 Context-free Grammars

Let $G = (V, P)$ be a rewriting system.

G is a **context free grammar**, $G = (N, \Sigma, P, S)$, if

$$N \cup \Sigma = V, N \cap \Sigma = \emptyset, S \in N, \text{ and}$$

$$P \subseteq N \times V^*.$$

A, B, C, \dots, S	nonterminal symbols	(N)
a, b, c, \dots, t	terminal symbols	(Σ)
X, Y, Z	general symbols	(V)
u, v, w, x, y, z	terminal strings	(Σ^*)
$\alpha, \beta, \gamma, \dots, \omega$	general strings	(V^*)
ε	empty string	

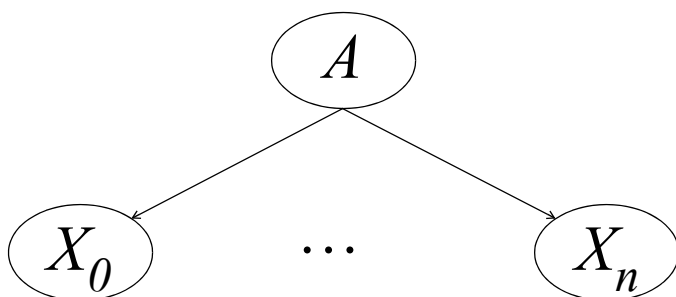
$$A \rightarrow \omega_1, \dots, A \rightarrow \omega_n.$$

$$A \rightarrow \omega_1 \mid \dots \mid \omega_n.$$

derivation tree (parse tree, syntax tree)

an labeled ordered tree.

$$A \rightarrow X_0 \dots X_n \in P, 0 \leq \forall i \leq n, X_i \in V$$



$$A < X_i, 0 \leq i \leq n$$

$$X_i < X_{i+1}, 0 \leq i < n$$

sentential forms of G

sentential forms of the start symbol S .

(sentential form of $\gamma \in V^$ was defined in 1.6)*

derivation of γ in G ($\gamma \in V^*$)

a derivation of string γ from S in G .

sentences of γ ($\gamma \in V^*$)

sentential forms of γ which are terminal strings

sentences of G

sentences of the start symbol S .

language generated by γ ($\gamma \in V^*$) in G

$$L_G(\gamma) = \{w \in \Sigma^* \mid \gamma \Rightarrow^* w \text{ in } G\}.$$

language generated (described) by G

$$L(G) = L_G(S) = \{w \in \Sigma^* \mid S \Rightarrow^* w \text{ in } G\}.$$

language L over Σ is context-free, if

$$\exists \text{ cfg } G = (N, \Sigma, P, S) \text{ such that } L(G) = L.$$

size of G

$$|G| = \sum |P|$$

norm of G

$$\|G\| = |G| \cdot \log_2 |V|$$

storing a grammar G in space $O(|G|)$

two linear lists, one for lhs, the other for rhs

4.2 Leftmost and Rightmost Derivation

Let $A \rightarrow \omega \in P$. Then

$$\Rightarrow_{lm} \equiv \{(xA\beta, x\omega\beta) \mid x \in \Sigma^*, \beta \in V^*\}.$$

$$\Rightarrow_{rm} \equiv \{(\alpha Ay, \alpha\omega y) \mid \alpha \in V^*, y \in \Sigma^*\}.$$

γ_1 **left(right)most derives** γ_2 in G using rule r

$$\gamma_1 \Rightarrow_{lm}^r \gamma_2, \quad (\gamma_1 \Rightarrow_{rm}^r \gamma_2) \quad \gamma_1, \gamma_2 \in V^*, r \in P.$$

γ_1 **left(right)most derives** γ_2 in G using rule string π

$$\gamma_1 \Rightarrow_{lm}^\pi \gamma_2, \quad (\gamma_1 \Rightarrow_{rm}^\pi \gamma_2) \quad \gamma_1, \gamma_2 \in V^*, \pi \in P^*.$$

$$\Rightarrow_{lm} = \bigcup_{r \in P} \Rightarrow_{lm}^r$$

$$\Rightarrow_{rm} = \bigcup_{r \in P} \Rightarrow_{rm}^r$$

γ_1 **directly left(right)most derives** γ_2 in G , if

$$\gamma_1 \Rightarrow_{lm} \gamma_2, \quad (\gamma_1 \Rightarrow_{rm} \gamma_2).$$

γ_1 **left(right)most derives** γ_2 in G , if

$$\gamma_1 \Rightarrow_{lm}^* \gamma_2, \quad (\gamma_1 \Rightarrow_{rm}^* \gamma_2).$$

γ_2 is a **left(right) sentential form** of γ_1 .

left(right)sentential form of G

left(right) sentential form of the start symbol S .

Let $n \geq 0$, $\gamma_i \in V^*$, $0 \leq i \leq n$. Then **left(right)most derivation** of length n from γ_0 to γ_n .

A string sequence $(\gamma_0, \dots, \gamma_n)$

$$\exists. 0 \leq \forall i \leq n, \gamma_i \Rightarrow_{lm} \gamma_{i+1} \ (\gamma_i \Rightarrow_{rm} \gamma_{i+1}).$$

$\gamma \Rightarrow_{lm}^{\pi} \gamma'$ ($\gamma \Rightarrow_{rm}^{\pi} \gamma'$), $\pi = r_1 \dots r_n$, $0 \leq \forall i \leq n$, $r_i \in P$, iff
 $0 \leq \forall i \leq n$, $\gamma_i \Rightarrow_{lm}^{r_i} \gamma_{i+1}$ ($\gamma_i \Rightarrow_{rm}^{r_i} \gamma_{i+1}$), $\gamma = \gamma_0$, $\gamma' = \gamma_n$.

left(right)most derivation of γ in G

left(right)most derivation of γ from S in G .

left(right) sentence of G

left(right) sentential forms of terminal string.

left sentence = right sentence

Lemma 4.1 Let $n \geq 0$, $m \geq 1$,

$$1 \leq \forall i \leq m, \alpha_i \in V^*, \beta \in V^*,$$

$\alpha_1 \dots \alpha_m \Rightarrow^n \beta$, $m \geq 1$, $n \geq 0$. Then

$\exists n_1, \dots, n_m \in \mathbb{N}$, $\beta_1, \dots, \beta_m \in V^*$ such that

$$\alpha_i \Rightarrow^{n_i} \beta_i, \sum_{i=1}^m n_i = n, \text{ and } \beta_1 \dots \beta_m = \beta.$$

Proof Induction on n

i) $n = 0$, $n_1 = \dots = n_m = 0$ and $\alpha_i = \beta_i$.

ii) $n > 0$, $A \rightarrow \omega \in P$

$$\alpha_1 \dots \alpha_m = \delta A \eta \Rightarrow \delta \omega \eta = \gamma_1 \dots \gamma_m \Rightarrow^{n-1} \beta.$$

$1 \leq \exists k \leq m$, $\exists \alpha_k = \mu A \nu$, $\gamma_k = \mu \omega \nu$; $\mu, \nu \in V^*$, and

$$1 \leq \forall i \neq k \leq m, \alpha_i = \gamma_i.$$

By IH $\exists n'_1, \dots, n'_m$, β_1, \dots, β_m such that

$$\gamma_i \Rightarrow^{n'_i} \beta_i, \sum_{i=1}^m n'_i = n-1, \text{ and } \beta_1 \dots \beta_m = \beta.$$

$$\alpha_k = \mu A \nu \Rightarrow \mu \omega \nu = \gamma_k.$$

$$\therefore n_i = n'_i, \text{ if } i \neq k; \text{ and } n_k = n'_k + 1$$

$$\therefore \alpha_i \Rightarrow^{n_i} \beta_i, \sum_{i=1}^m n_i = n.$$

Theorem 4.2 Let γ be a general string ($\gamma \in V^*$). Then $\{w \in \Sigma^* \mid \gamma \Rightarrow_{lm}^n w\} = \{w \in \Sigma^* \mid \gamma \Rightarrow^n w\} = \{w \in \Sigma^* \mid \gamma \Rightarrow_{rm}^n w\}$

Proof

$\{w \in \Sigma^* \mid \gamma \Rightarrow_{lm}^n w\} \subseteq \{w \in \Sigma^* \mid \gamma \Rightarrow^n w\}$ is trivial,
since $\Rightarrow_{lm} \subseteq \Rightarrow$.

1. prove $X \Rightarrow^n u$ implies $X \Rightarrow_{lm}^n u$.

i) if $n=0$, it is trivial

ii) if $n>0$

$$\exists X \Rightarrow X_1 \dots X_m \Rightarrow^{n-1} u$$

if $m=0$ it is trivial ($u=\varepsilon$)

if $m>0$, $\exists m \geq 1: u = u_1 \dots u_m, X_i \Rightarrow^{n_i} u_i, \sum^m n_i = n-1$. **L4.1**

Furthermore $X_i \Rightarrow_{lm}^{n_i} u_i$ by IH

$$\therefore X \Rightarrow_{lm} X_1 \dots X_m \Rightarrow_{lm}^{n-1} u.$$

2. prove $\gamma = Y_1 \dots Y_k \Rightarrow^n w$ implies $\gamma \Rightarrow_{lm}^n w, Y_i \in V$.

$Y_i \Rightarrow^{n_i} w_i, \sum^k n_i = n$ by Lemma 4.1.

$Y_i \Rightarrow_{lm}^{n_i} w_i$ by the above condition 1.

$$\therefore \gamma = Y_1 Y_2 \dots Y_k \Rightarrow_{lm}^{n_1} w_1 Y_2 \dots Y_k \Rightarrow_{lm}^{n_2} \dots \Rightarrow_{lm}^{n_k} w_1 w_2 \dots w_k$$

$$\therefore \gamma \Rightarrow_{lm}^n w \therefore \{w \in \Sigma^* \mid \gamma \Rightarrow_{lm}^n w\} \supseteq \{w \in \Sigma^* \mid \gamma \Rightarrow^n w\}$$

Furthermore $\{w \in \Sigma^* \mid \gamma \Rightarrow^n w\} = \{w \in \Sigma^* \mid \gamma \Rightarrow_{rm}^n w\}$ is dual

Corollary 4.3 $L(G)$ coincides with the set of left sentences and right sentences of G .

4.3 Ambiguity of Grammars

A grammar is **ambiguous**, iff some sentence has **more than one leftmost derivation**.

$$S \rightarrow a \mid \text{ifcl } S \mid \text{ifcl } S \text{ else } S$$

$$E \rightarrow a \mid E + E \mid E * E$$

$$S \rightarrow a \mid S \quad \text{infinite ambiguity}$$

A context-free language is **inherently ambiguous**, if every grammar is ambiguous.

Proposition 4.4(Parikh (1966)) There exist inherently ambiguous language; one example is the language $\{a^i b^j c^k \mid i=j \text{ or } j=k\}$.

ambiguity of grammar

leftmost derivation definition

rightmost derivation theorem(proof)

Fact 4.5 Let $\pi \in P^*$.

(1) if $\alpha \Rightarrow_{lm}^{\pi} \beta$ and $\alpha \Rightarrow_{lm}^{\pi} \beta'$, then $\beta = \beta'$.

(2) if $\alpha \Rightarrow_{rm}^{\pi} \beta$ and $\alpha \Rightarrow_{rm}^{\pi} \beta'$, then $\beta = \beta'$.

same rule string \Rightarrow same sentential form

Lemma 4.6 Let $m \geq 1$, $\pi \in P^*$ and $\alpha_1 \dots \alpha_m \Rightarrow_{lm}^{\pi} w \in \Sigma^*$.

Then \exists **unique** π_1, \dots, π_m and w_1, \dots, w_m such that

$$\pi = \pi_1 \dots \pi_m, w = w_1 \dots w_m, \alpha_i \Rightarrow_{lm}^{\pi_i} w_i \quad 1 \leq i \leq m.$$

Proof

The existence of such strings are trivial. (**L4.1**)

Then we should prove the uniqueness.

$$\pi_1 \dots \pi_m = \pi'_1 \dots \pi'_m = \pi,$$

$$w_1 \dots w_m = w'_1 \dots w'_m = w,$$

$$\alpha_i \Rightarrow_{lm}^{\pi_i} w_i, \text{ and } \alpha_i \Rightarrow_{lm}^{\pi'_i} w'_i \text{ for } 1 \leq i \leq m.$$

show that $\pi_i = \pi'_i$ and $w_i = w'_i$ for $1 \leq i \leq m$.

Induction on $m \geq 1$. ($m=1$, trivial)

induction hypothesis

$$\pi_1 \dots \pi_{j-1} = \pi'_1 \dots \pi'_{j-1} \text{ and } w_1 \dots w_{j-1} = w'_1 \dots w'_{j-1}.$$

π_j is a prefix of π'_j ; or vice versa

assume w.o.l.g $\pi'_j = \pi_j \pi''_j$

$$\alpha_j \Rightarrow_{lm}^{\pi'_j} w'_j \text{ and } \alpha_j \Rightarrow_{lm}^{\pi_j} w_j \therefore \alpha_j \Rightarrow_{lm}^{\pi_j} w_j \Rightarrow_{lm}^{\pi''_j} w''_j$$

$$w'_j = w''_j \text{ by Fact 4.5, and since } w_j \in \Sigma^*, w_j = w''_j.$$

$$\text{so, } \pi''_j = \varepsilon, \pi_j = \pi'_j, \text{ and } w_j = w'_j.$$

Lemma 4.7 Let $m \geq 1$ and $\alpha_1 \dots \alpha_m \Rightarrow_{rm}^{\pi} w \in \Sigma^*$. Then

\exists **unique** π_1, \dots, π_m and w_1, \dots, w_m such that

$$\pi = \pi_m \dots \pi_1, w = w_1 \dots w_m, \alpha_i \Rightarrow_{rm}^{\pi_i} w_i \quad 1 \leq i \leq m.$$

Lemma 4.8 Let $A \in N$, $\pi \in P^+$. Then

$$A \Rightarrow^r X_1 \dots X_m \Rightarrow_{lm}^{\pi'} w \text{ where } \pi = r\pi', r \in P, \pi' \in P^*.$$

Moreover there exist a **unique** strings

$$\pi'_1 \dots \pi'_m = \pi', w_1 \dots w_m = w, \text{ such that}$$

$$X_i \Rightarrow_{lm}^{\pi'_i} w_i \text{ for } 1 \leq i \leq m.$$

Define a function f : leftmost deri. \rightarrow rightmost deri.

$$f: P^* \rightarrow P^*.$$

$$(1) f(\varepsilon) = \varepsilon.$$

$$(2) f(\pi) = f(r \cdot \pi_1 \dots \pi_m) = r \cdot f(\pi_m) \dots f(\pi_1),$$

if $r = A \rightarrow X_1 \dots X_m$, $\pi = r \cdot \pi_1 \dots \pi_m$, and

$$\pi_i \in P^* \text{ for } 1 \leq i \leq m (\pi \in P^+)$$

Lemma 4.9 Let $X \in V \cup \{\varepsilon\}$.

If $X \Rightarrow_{lm}^{\pi} w \in \Sigma^*$ then $X \Rightarrow_{rm}^{f(\pi)} w$.

Proof

i) $\pi = \varepsilon$, trivial ($f(\varepsilon) = \varepsilon$).

ii) $|\pi| > 0$, $\pi = r\pi'$, $X \Rightarrow_{lm}^r X_1 \dots X_m \Rightarrow_{lm}^{\pi'} w$

if $\pi' = \varepsilon$, $X \Rightarrow_{rm}^{f(r)} w$, since $f(r) = r$.

if $\pi' \neq \varepsilon$, $\pi'_1 \dots \pi'_m = \pi'$, $w_1 \dots w_m = w$, $X_i \Rightarrow_{lm}^{\pi'_i} w_i$ (**L4.6**)

by IH, $X_i \Rightarrow_{rm}^{f(\pi'_i)} w_i$, and $X \Rightarrow_{rm}^r X_1 \dots X_m \Rightarrow_{rm}^{f(\pi'_m) \dots f(\pi'_1)} w$

$\therefore X \Rightarrow_{rm}^{rf(\pi'_m) \dots f(\pi'_1)} w (= \Rightarrow_{rm}^{f(\pi)} w)$.

Let $w \in \Sigma^*$. For any $X \in V \cup \{\varepsilon\}$,

$$\Pi_L(X) = \{\pi \mid X \Rightarrow_{lm}^{\pi} w\}.$$

Let $\Pi_L = \cup_{X \in V \cup \{\varepsilon\}} \Pi_L(X)$.

$$\Pi_R(X) = \{\pi \mid X \Rightarrow_{rm}^{\pi} w\}.$$

$$\Pi_R = \cup_{X \in V \cup \{\varepsilon\}} \Pi_R(X).$$

$f: \Pi_L(\subseteq P^*) \rightarrow \Pi_R(\subseteq P^*)$.

Lemma 4.10 Let $X \in V \cup \{\varepsilon\}$. If $X \Rightarrow_{rm}^{\pi} w \in \Sigma^*$ then

$$X \Rightarrow_{lm}^{\pi'} w \text{ where } \pi = f(\pi').$$

f is onto.

Lemma 4.11 Let $X \in V \cup \{\varepsilon\}$. If $X \Rightarrow_{rm}^{\pi} w \in \Sigma^*$ then

$$X \Rightarrow_{lm}^{\pi'} w, X \Rightarrow_{lm}^{\pi''} w, \text{ and } f(\pi') = f(\pi'') = \pi,$$

implies that $\pi' = \pi''$.

f is one-to-one.

$\therefore f$ is one-to-one onto.

Since f is a bijection,
 one-to-one correspondence
 between left and right parses

Theorem 4.12 Let $G = (N, \Sigma, P, S)$ be a grammar. G is ambiguous iff some sentence in G has **more than one rightmost derivations**.

Proof

→) Assume G is ambiguous, then $w \in L(G)$ has at least two leftmost derivation.

$$S \Rightarrow_{lm}^{\pi} w \text{ and } S \Rightarrow_{lm}^{\pi'} w$$

$$S \Rightarrow_{rm}^{f(\pi)} w \text{ and } S \Rightarrow_{rm}^{f(\pi')} w$$

by Lemma 4.11, $f(\pi) \neq f(\pi')$ because $\pi \neq \pi'$

$\therefore w$ has more than one rightmost derivation

←) $w \in L(G)$ has at least two rightmost derivation

$$S \Rightarrow_{rm}^{\pi} w \text{ and } S \Rightarrow_{rm}^{\pi'} w$$

by Lemma 4.10, $S \Rightarrow_{lm}^{f^{-1}(\pi)} w$ and $S \Rightarrow_{lm}^{f^{-1}(\pi')} w$

π and π' are distinct, so $f^{-1}(\pi)$ and $f^{-1}(\pi')$ are distinct

$\therefore G$ is ambiguous.

4.4 Useless and Nullable Symbols

A symbol $X \in V$ is **useful** if

$$S \Rightarrow^* \alpha X \beta \Rightarrow^* w, \alpha \beta \in V^*, w \in \Sigma^*.$$

Otherwise **useless**.

A grammar G is **reduced**, if

it contains **no useless symbols**.

A nonterminal $A \in N$ is **nullable**, if $A \Rightarrow^+ \varepsilon$ in G .

A rule of the form $A \rightarrow \varepsilon$ is called an **ε -rule**.

A grammar having no ε -rule is called **ε -free**.

$$W = \{A \mid A \Rightarrow^* \varepsilon\}$$

$$W_1 = \{A \mid A \rightarrow \varepsilon \in P\}$$

$$W_k = \{A \mid A \notin \bigcup_{i=1}^{k-1} W_i, A \rightarrow A_1 \dots A_n \in P, \forall A_j \in \bigcup_{i=1}^{k-1} W_i\}$$

$$(\exists j . \exists. A_j \in W_{k-1})$$

Lemma 4.13 $W = \bigcup_{i=1}^{|W|} W_i$.

Proof

$$W = \bigcup_{i=1}^{\infty} W_i$$

if $A \in W_k$, $k > 1$, then $A \notin \bigcup_{i=1}^{k-1} W_i$

$\exists A \rightarrow A_1 \dots A_n$ such that $\forall A_j \in \bigcup_{i=1}^{k-1} W_i$

some A_j must be in W_{k-1} , thus $W_{k-1} \neq \emptyset$ if $W_k \neq \emptyset$

Theorem 4.14 The set of nullable nonterminals of a grammar G can be computed in time $(O|G|)$.

$W_1 := \{A \mid A \rightarrow \varepsilon \in P\};$
 $k := 1;$
while $W_k \neq \emptyset$ **do**
 mark A 's in W_k ;
 $W_{k+1} := \{A \mid A \rightarrow A_1 \dots A_m \in P, \forall A_i: \text{marked},$
 $A: \text{not marked}\};$
 $k := k + 1$
 od;
 $W = \emptyset;$
for $i:=1$ **to** $k-1$ **do** $W := W \cup W_k$ **od**.

Lemma 4.15 Let $G = (N, \Sigma, P, S)$ and $G' = (N, \emptyset, P', S)$ where

$$P' = \{A \rightarrow \alpha_1 \dots \alpha_n \mid A \rightarrow x_0 \alpha_1 x_1 \dots \alpha_n x_n \in P, \\ 1 \leq \forall i \leq n, \alpha_i \in N^+, 0 \leq \forall j \leq n, x_j \in \Sigma^*\}.$$

Then $A \Rightarrow^* x \in \Sigma^*$ in G , iff $A \Rightarrow^* \varepsilon$ in G' .

Let *contains* be a relation on V defined by

A *contains* X if $A \rightarrow \alpha X \beta \in P$, $A \in N$, $X \in V$.

Lemma 4.16 S *contains*^{*} X , iff $S \Rightarrow^* \alpha X \beta$, $\alpha, \beta \in V^*$.

Proof If $X \in \text{contains}^n(S)$, $S \Rightarrow^n \alpha X \beta$ for $n \geq 0$.

If $S \Rightarrow^n \alpha X \beta$, $X \in \text{contains}^m(S)$, for $0 \leq m \leq n$.

Theorem 4.17 Any grammar G can be transformed into an equivalent **reduced** grammar in time $O(|G|)$.

Guarded Commands, Nondeterminancy, and Formal Derivation of Programs

E.W. Dijkstra, CACM 18,8 pp.453-457, (Aug. 75)

A Discipline of Programming, Prentice-Hall, 1976

Concurrent assignment

$x, y := y, x$

$x := y; y := x;$

$y := x; x := y$

$t := x; x := y; y := t$

$x_1, \dots, x_n := E_1, \dots, E_n.$

Nondeterminancy

if $x \geq y \rightarrow m := x$

/ $x \leq y \rightarrow m := y$

fi

If $x=y, m := x$ ***or*** $m := y$

Repeatative construct

$i, S := 1, 0; \mathbf{do} \ i \leq 100 \rightarrow S := S + i; i := i + 1 \mathbf{od}$

$1 \leq \forall i \leq 101: S = \sum_{j=1}^{i-1} j$

If $i = 101, S = \sum_{j=0}^{100} j$

$i, S := 0, 0; \mathbf{do} \ i \leq 99 \rightarrow i := i + 1; S := S + i \mathbf{od}$

$0 \leq \forall i \leq 100: S = \sum_{j=0}^i j$

If $i = 100, S = \sum_{j=0}^{100} j$

$q_1, q_2, q_3, q_4 := Q_1, Q_2, Q_3, Q_4;$
do $q_1 > q_2 \rightarrow q_1, q_2 := q_2, q_1$
 | $q_2 > q_3 \rightarrow q_2, q_3 := q_3, q_2$
 | $q_3 > q_4 \rightarrow q_3, q_4 := q_4, q_3$
od.

$$\begin{aligned}
 P &= 1 \leq \forall i \leq 4: q_i \text{'s are permutation of } Q_i \\
 \neg BB &= \neg((q_1 > q_2) \vee (q_2 > q_3) \vee (q_3 > q_4)) \\
 &= \neg(q_1 > q_2) \wedge \neg(q_2 > q_3) \wedge \neg(q_3 > q_4) \\
 &= (q_1 \leq q_2) \wedge (q_2 \leq q_3) \wedge (q_3 \leq q_4) \\
 &= (q_1 \leq q_2 \leq q_3 \leq q_4)
 \end{aligned}$$

P *Loop invariance condition*

BB *There exists at least one guard that is true*

*If P is true before the loop(initialization of P),
 P remains true in the loop(loop invariance), and
 the loop terminates(loop terminating),
 then $P \wedge \neg BB$ is true after the loop.*

if $X > 0$ **and** $Y > 0 \rightarrow$

$x, y := X, Y;$

do $x > y \rightarrow x := x - y$

| $x < y \rightarrow y := y - x$ **od fi**; *print*(x).

Theorem 4.18 Let $G = (N, \Sigma, P, S)$ be a grammar and $m \geq 1$ is the length of the longest rule in P . Then G can be transformed in time $O(2^m \cdot |G|)$ into a grammar $\hat{G} = (N, \Sigma, \hat{P}, S)$ satisfying

- (1) \hat{G} is ε -free.
- (2) $L_{\hat{G}}(X) = L_G(X) \setminus \{\varepsilon\}$ for all $X \in V$.

Proof

$$\hat{P} = \{A \rightarrow \alpha_1 \dots \alpha_n \mid \exists A_1, \dots, A_{n-1}. \exists. 1 \leq \forall i < n, A_i \Rightarrow^* \varepsilon, \\ A \rightarrow \alpha_1 A_1 \dots \alpha_{n-1} A_{n-1} \alpha_n \in P, n \geq 1, \alpha_1 \dots \alpha_n \neq \varepsilon\}$$

If $\alpha \Rightarrow_{\hat{G}} \beta$, then $\alpha \Rightarrow_G^+ \beta$. $\therefore L_{\hat{G}}(X) \subseteq L_G(X)$.

If $X \Rightarrow_G^n w$, then $X \Rightarrow_{\hat{G}}^* w$. $X \in V$ and $w \in \Sigma^+$.

assume $X \Rightarrow_G^{n'} w$ implies $X \Rightarrow_{\hat{G}}^* w$ when $n' < n$.

$$X \Rightarrow_G X_1 \dots X_k \Rightarrow_G^{n-1} w \in \Sigma^+.$$

$$n-1 = \sum^k n_i, w = w_1 \dots w_k, X_i \Rightarrow_G^{n_i} w_i, 1 \leq i \leq k \text{ (L4.6)}$$

Let $\{i_1, \dots, i_l\} \subseteq \{1, \dots, k\} (1 \leq \forall m < l, i_m < i_{m+1})$

where $w_{i_1} \dots w_{i_l} = w$ and $(1 \leq \forall j < l, w_{i_j} \neq \varepsilon)$

by IH, $X_{i_j} \Rightarrow_{\hat{G}}^* w_{i_j}$, $X_{i_1} \dots X_{i_l} \Rightarrow_{\hat{G}}^* w_{i_1} \dots w_{i_l} = w$

by construction, \hat{G} has a rule $X \rightarrow X_{i_1} \dots X_{i_l}$

$$X \Rightarrow_{\hat{G}}^* w, \therefore L_G(X) \setminus \{\varepsilon\} \subseteq L_{\hat{G}}(X)$$

\hat{G} has at most 2^k rules for $A \rightarrow X_1 \dots X_k$

The transform requires $O(2^m \cdot |G|)$ time.

4.5 Canonical Two-Form Grammar

A Grammar $G = (N, \Sigma, P, S)$ is in **canonical two-form** if its rules are of the forms

$$A \rightarrow BC, \quad A \rightarrow B, \quad A \rightarrow a, \quad S \rightarrow \varepsilon.$$

Furthermore, if $S \rightarrow \varepsilon \in P$, S may not occur in the right-hand side of any rule.

Lemma 4.19 Any Grammar $G = (N, \Sigma, P, S)$ can be transformed in Time $O(|G|)$ into $G' = (N', \Sigma, P', S')$ satisfying

(1) P' are of the forms

$$A \rightarrow BC, \quad A \rightarrow B, \quad A \rightarrow a, \quad A \rightarrow \varepsilon.$$

(2) $N \subseteq N'$

(3) $L_{G'}(X) = L_G(X)$ for all $X \in V$

Proof

Let $A \rightarrow X_1 \dots X_m \in P$, $m \geq 2$. Then

$$[A] \rightarrow [X_1][X_2 \dots X_m],$$

$$[X_2 \dots X_m] \rightarrow [X_2][X_3 \dots X_m],$$

...

$$[X_{m-1}X_m] \rightarrow [X_{m-1}][X_m],$$

and

$$[X_i] \rightarrow X_i \quad 1 \leq i \leq m.$$

$$N' = N \cup \{[X\beta] \mid A \rightarrow \alpha X\beta \in P, \alpha, \beta \in V^+\} \\ \cup \{[X] \mid X \in V, A \rightarrow \alpha X\beta \in P\}.$$

$$\begin{aligned}
P' = & \{A \rightarrow \alpha \mid A \rightarrow \alpha \in P, |\alpha| \leq 1\} \\
& \cup \{[A] \rightarrow [X][\beta] \mid A \rightarrow X\beta \in P, \beta \in V^+\} \\
& \cup \{[X\beta] \rightarrow [X][\beta] \mid A \rightarrow \alpha X\beta \in P, \alpha, \beta \in V^+\} \\
& \cup \{[X] \rightarrow X \mid A \rightarrow \alpha X\beta \in P\}.
\end{aligned}$$

i) $L_G(X) \subseteq L_{G'}(X)$ is trivial, since

$$\alpha \Rightarrow_G \beta \text{ implies } \alpha \Rightarrow_G^+ \beta.$$

ii) $L_{G'}(X) \subseteq L_G(X)$.

(case1) $X \Rightarrow_G^n w$ implies $X \Rightarrow_G^* w$, and

(case2) $[\alpha] \Rightarrow_G^n w$ implies $\alpha \Rightarrow_G^* w$.

The case $n=0$ is trivial.

IH: Assume that $n > 0$ and it holds for less than n .

case 1) Consider $X \Rightarrow_G^n w$, since $n > 0$,

$$X \Rightarrow_G \eta \Rightarrow_{G'}^{n-1} w.$$

if $\eta = \varepsilon$, $X \rightarrow \varepsilon \in P$, by construction

if $|\eta| = 1$, $\eta = Y$

$Y \Rightarrow_G^* w$ (IH case 1) and $X \rightarrow Y \in P$ (construction)

$$\therefore X \Rightarrow_G^* w.$$

if $|\eta| > 1$, $\eta = [Y][\beta]$

$X \rightarrow Y\beta \in P$, by construction

$Y\beta \Rightarrow_G^* w$, by Lemma 4.6 and IH case 2

$$\therefore X \Rightarrow_G^* w.$$

case 2) Consider $[\alpha] \Rightarrow_G^n w$. Then,

$$[\alpha] \Rightarrow_G, \eta \Rightarrow_G^{n-1} w, \text{ for some } \eta \in V'^*$$

if $|\alpha|=1$, $\eta=\alpha$ and by IH case 1

$$\therefore \alpha \Rightarrow_G^* w$$

if $\alpha=X\gamma$, $X \in V$ and $\gamma \in V^+$

$$[X\gamma] \Rightarrow_G, [X][\gamma] \Rightarrow_G^{n-1} w, \text{ by construction}$$

$$X \Rightarrow_G^* w_1 \text{ and } \gamma \Rightarrow_G^* w_2, \text{ where } w=w_1w_2$$

by Lemma 4.6 and IH case 2

$$\therefore X\gamma \Rightarrow_G^* w, \quad \text{by Lemma 4.6}$$

$\forall A \rightarrow X_1 \dots X_m \in P$ of G of length $m+1$

G' has rules of length $3(m-1) + 2m = 5m - 3$.

\therefore Construction of G' is $O(|G|)$.

Theorem 4.20 Any G can be transformed into G'

(1) G' is **canonical two-form**,

(2) $N \subseteq N'$,

(3) $L_{G'}(X) = L_G(X) \setminus \{\epsilon\} \quad \forall X \in V$,

(4) $L_{G'}(S) = L_G(S)$

Proof $G_1 = (N_1, \Sigma, P_1, S)$ in Lemma 4.19 in $O(|G|)$

Find nullable nonterminal of G_1 in $O(|G|)$ (**T4.14**)

Transform G_1 into ε -free $G_2 = (N_1, \Sigma, P_2, S)$,

$$L_{G_2}(X) = L_{G_1}(X) \setminus \{\varepsilon\}, \quad \forall X \in V_1.$$

$$O(2^m \cdot |G_1|) = O(2^2 \cdot |G_1|) = O(|G|), \quad (\mathbf{T4.18})$$

If S is not nullable condition (4) holds,
otherwise,

$$G' = (N_1 \cup \{S'\}, \Sigma, P_2 \cup \{S' \rightarrow S, S' \rightarrow \varepsilon\}, S').$$

A grammar G is **Chomsky normal form**, if P are

$$A \rightarrow BC, A \rightarrow a, \text{ and } S \rightarrow \varepsilon.$$

Canonical two-form \rightarrow Chomsky normal form

Elimination rule of form

$$A \rightarrow B \quad \text{unit rules (single rules)}$$

$$\forall A \rightarrow BC, B \Rightarrow^* B', C \Rightarrow^* C'.$$

$$A \rightarrow B'C'$$

eliminating all rules of the form $A \rightarrow B$

$$O(|G|) \text{ for } B \Rightarrow^* B'$$

$$O(|N|^2) \text{ for } A \rightarrow B'C'$$

$$O(|G| \cdot |N|^2)$$

No better algorithm is known.

4.6 Derivational Complexity

$$TIME_G(w) = TIME_G(S, w)$$

$$SPACE_G(w) = SPACE_G(S, w)$$

$$O(|w|)$$

Consider an ε -free grammar G_n with rules

$$A_1 \rightarrow A_2,$$

$$A_2 \rightarrow A_3,$$

...

$$A_n \rightarrow a \mid A_1 A_1$$

$$L(A_1) = a^+,$$

$$TIME(a^k) = 2nk - n$$

$$SPACE(a^k) = k$$

Theorem 4.21 An ε -free grammar G , $|N|=n$.

If $X \in V$, $w \in L(X)$, then $\text{TIME}(X, w) = 2n|w| - n$ and $\text{SPACE}(X, w) = |w|$. And, these bounds are minimal.

Proof

i) $\text{SPACE}(X, w) = |w|$, because no step decreases the length of the sentential form in ε -free grammar

ii) $\text{TIME}(X, w) = 2n|w| - n$

by induction on $|w|$

if $|w|=1$, $X \Rightarrow^\pi w$, π contains only $A \rightarrow B$ or $A \rightarrow a$
 $|\pi| \leq n = 2n|w| - n$, because two rules with same lhs in π make unnecessary loop.

if $|w| > 1$, $X \Rightarrow^* w$

$$\exists r = A \rightarrow X_1 \dots X_m$$

$$X \Rightarrow^{\pi'} A \Rightarrow^r X_1 \dots X_m \Rightarrow^\pi w$$

$$\pi = \pi_1 \dots \pi_m, w = w_1 \dots w_m, X_i \Rightarrow^{\pi_i} w_i \quad (\mathbf{L4.6})$$

π' contains only the form $A \rightarrow B$, so $|\pi'| \leq n-1$

$|\pi_i| \leq 2n|w_i| - n$, by IH

$$\begin{aligned} |\pi' r \pi| &= |\pi'| + 1 + \sum_{i=1}^m |\pi_i| \\ &\leq (n-1) + 1 + \sum_{i=1}^m (2n|w_i| - n) \\ &= 2n|w| - (1-m)n \\ &\leq 2n|w| - n \end{aligned}$$

The grammar G_n , $n \geq 1$, show this bound is minimal

Consider a grammar $G_{n,m}$, $n \geq 1$, $m \geq 2$ with rules

$$A_1 \rightarrow A_2^m,$$

$$A_2 \rightarrow A_3^m,$$

...

$$A_{n-1} \rightarrow A_n^m,$$

$$A_n \rightarrow \varepsilon$$

$$A_i \Rightarrow^1 A_{i+1}^m \Rightarrow^m A_{i+2}^{m^2} \Rightarrow^{m^2} \dots \Rightarrow^{m^{n-i-1}} A_n^{m^{n-i}} \Rightarrow^{m^{n-i}} \varepsilon$$

$$TIME(A_i, \varepsilon) = 1 + m + \dots + m^{n-i} = (m^{n-i+1} - 1) / (m - 1)$$

$$A_i \Rightarrow_{lm}^* A_n^m A_{n-1}^{m-1} \dots A_{i+1}^{m-1}$$

$$\Rightarrow_{lm}^* A_n^m A_{n-1}^{m-2} \dots A_{i+1}^{m-1}$$

...

$$\Rightarrow_{lm}^* A_n^m A_{n-1}^{m-1} A_{n-2}^{m-2} A_{n-3}^{m-1} \dots A_{i+1}^{m-1}$$

...

$$\Rightarrow_{lm}^* A_n^m A_{n-1} A_{n-2} \dots A_{i+1}$$

$$\Rightarrow_{lm}^* A_{n-1} A_{n-2} \dots A_{i+1}$$

...

$$\Rightarrow_{lm}^* \varepsilon$$

$$SPACE(A_i, \varepsilon) = (n - i)(m - 1) + 1$$

$$\therefore TIME(A_i, \varepsilon) = (m^{k_i} - 1) / (m - 1)$$

$$SPACE(A_i, \varepsilon) = (k_i - 1)(m - 1) + 1$$

where $k_i = n - i + 1$.

$$W_k = \{A \mid \exists A \Rightarrow^k \varepsilon, \text{ but } \nexists n \leq k. A \Rightarrow^n \varepsilon, n \leq k\}$$

Lemma 4.22 Let G be a grammar and $m \geq 2$ the length of the rhs of the longest rule. Then $\forall k$ and

$$A \in W_k \quad \text{TIME}(A, \varepsilon) = (m^k - 1)/(m - 1) \text{ and}$$

$$\text{SPACE}(A, \varepsilon) = (k - 1)(m - 1) + 1.$$

And, these bounds are minimal.

Proof induction on k .

If $k = 1$ and $A \in W_k$, $A \rightarrow \varepsilon \in P$.

$$\therefore \text{time } 1 = (m^1 - 1)/(m - 1)$$

$$\text{space } 1 = (1 - 1)(m - 1) + 1.$$

Assume $k > 1$, $\forall k' < k$, $A' \in W_{k'}$,

$$\text{time } (m^{k'} - 1)/(m - 1), \text{ space } (k' - 1)(m - 1) + 1.$$

If $A \in W_k$, $r = A \rightarrow A_1 \dots A_l \in P$, $\forall A_i \in W_{k_i}$, $k_i < k$,

$$A_i \Rightarrow^{\pi_i} \varepsilon, |\pi_i| \leq (m^{k_i} - 1)/(m - 1).$$

$$\text{SPACE}(A_i, \varepsilon) = (k_i - 1)(m - 1) + 1,$$

$$\therefore \text{TIME}(A, \varepsilon) \leq 1 + \sum_{i=1}^l (m^{k_i} - 1)/(m - 1)$$

$$\leq 1 + l(m^{k-1} - 1)/(m - 1)$$

$$\leq 1 + m(m^{k-1} - 1)/(m - 1)$$

$$= (m^k - 1)/(m - 1).$$

4.7 Context-free Language Recognition

$P_{mem}(G)$: Given a context-free grammar G in the class \mathcal{G} of all such grammars, and a string w , is $w \in L(G)$?

Theorem 4.27 given cfg $G=(N, \Sigma, P, S)$ and $w \in \Sigma^*$, $w \in L(G)$?

$NTIME(O(|G| \cdot |w|))$ and $NSPACE(O(|G| + |w|))$

Proof

assume G is in canonical two-form

guess a derivation $(\gamma_0, \dots, \gamma_n)$, $\gamma_0 = S$, $\gamma_n \in \Sigma^*$, $\gamma_n = w$?

$w = a_1 \dots a_n$, $N(i, j) = \{A \mid A \text{ derives } a_i \dots a_j\}$, $1 \leq i \leq j \leq n$

B unit-rule X if $B \rightarrow X \in P$

Fact 4.28 $N(i, i) = (\text{unit-rule}^{-1})^+(a_i)$

Lemma 4.29 for all i, j , $1 \leq i \leq j \leq n$,

$$N(i, j) = \bigcup_{k=0}^{j-i-1} (\text{unit-rule}^{-1})^*(\{A \mid A \rightarrow BC \in P,$$

$$B \in N(i, i+k), C \in N(i+k+1, j)\})$$

if $S \in N(1, n)$, then $a_1 \dots a_n \in L_G(S)$

Theorem 4.30 given cfg $G=(N, \Sigma, P, S)$ and $w \in \Sigma^*$, $w \in L(G)$?

$DTIME(O(|G| \cdot |w|^3))$ and $DSPACE(O(|G| \cdot |w|^2))$